R E P O R T R E S U M E S

ED 003 060 24

SAMPLING DISTRIBUTIONS OF ERROR IN MULTIDIMENSIONAL SCALING.

BY- STAKE, ROBERT E. AND OTHERS
NEBRASKA UNIV., LINCOLN

REPORT NUMBER CRP-1253 PUB DATE JUN 62
EDRS PRICE MF-$0.18 HC-$3.12 78P.

DESCRIPTORS- *STATISTICAL ANALYSIS, DATA ANALYSIS, *RESEARCH METHODOLOGY, *MATHEMATICAL MODELS, *TEST CONSTRUCTION, MEASUREMENT, EVALUATION, LINCOLN, NEBRASKA

AN EMPIRICAL STUDY WAS MADE OF THE ERROR FACTORS IN MULTIDIMENSIONAL SCALING (MDS) TO REFINE THE USE OF MDS FOR MORE EXPERT MANIPULATION OF SCALES USED IN EDUCATIONAL MEASUREMENT. THE PURPOSE OF THE RESEARCH WAS TO GENERATE TABLES OF THE SAMPLING DISTRIBUTIONS THAT ARE NECESSARY FOR DISCRIMINATING BETWEEN ERROR AND NONERROR MDS DIMENSIONS. THE TABLES DEVELOPED WERE USED TO TEST THE HYPOTHESIS THAT A GIVEN MDS FACTOR IS NO GREATER THAN ONE ATTRIBUTABLE TO A COMBINATION OF MEASUREMENT ERRORS. SEVERAL VARIABLES THAT COULD INFLUENCE SAMPLING DISTRIBUTIONS OF ERROR ROOTS WERE CONSIDERED. THIS RESEARCH INVOLVED THE SIMULATION OF JUDGEMENTS OF 15 OBJECTS BY 35 JUDGES ACCORDING TO THE MULTIDIMENSIONAL RANKING PROCEDURE OF KLINGBERG (1941). THE EFFECTS OF THE VARIABLES WERE CONSIDERED BY VARYING THE NUMBER OF JUDGES FROM 20 TO 80 AND BY INDEPENDENTLY VARYING THE NUMBER OF OBJECTS FROM 8 TO 15. COMPUTER PROGRAMS FOR THIS INVESTIGATION WERE BASED ON PROCEDURES OUTLINED BY TORGERSON (1958). THE STUDY LED TO THE CONCLUSIONS THAT (1) ONLY A SINGLE NONERROR DIMENSION COULD BE DESIGNATED WITH STATISTICAL CONFIDENCE, AND (2) ASCERTAINING THE NONERROR DIMENSIONALITY OF THE STRUCTURE COULD NOT BE ACCOMPLISHED WITH THESE STRUCTURE STATISTICS ALONE. (AH)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

SAMPLING DISTRIBUTIONS OF ERROR IN MULTIDIMENSIONAL SCALING

Cooperative Research Project No. 1253

Robert E. Stake

James H. Beaird

Douglas D. Sjogren

University of Nebraska

June, 1962

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ED003060

3814
179

ACKNOWLEDGMENTS

The authors wish to acknowledge the advice and assistance provided by many individuals in the planning and the execution of this project. Consultation on theoretical consideration came from Ledyard Tucker, Samuel Messick, Roger Pinkham, and Bernard Harris. Carl Helm provided a computer program for generating random numbers, and John Christopher assisted in the development of the total computer program. Editorial advice was received from Jonathan Warren.

Graduate students involved in this project at various stages were Don Anderson, Jaswant Bedi, Ronald Ellis, Gene Glass, and William Walker.

The figures were drawn by Vernon Hines. Various manuscripts were typed by Nickole Christie, Agnes Morris, and Carolyn Orr.

This research study was made possible by support from the Cooperative Research Program of the U. S. Office of Education.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	MULTIDIMENSIONAL SCALING	1
	Objectives of This Study	2
	Theoretical Background	3
	Young and Householder's Solution	5
	Torgerson's B* Matrix	7
	Torgerson's MDS Solution	7
	Estimating the Additive Constant	8
	Factor Analysis	9
	Procedure	12
	Phase II	13
II.	STRUCTURE STATISTICS	14
	A General Solution for Characteristic Roots	14
	Empirical Distributions	15
	Alpha	17
	Beta	19
	Gamma	24
	Delta	26
	Epsilon	29
	The Additive Constant	34
III.	ARTIFICIAL STRUCTURE	37
	Scale Values	38
	The Sort	39
	The Structure Statistics and AIPD	39
	Comparison of Statistics as to Sensitivity	39
	Additional Dimensionality	46
	Evaluating the Lesser Dimensions	49
	Increased Error Variance	50

CHAPTER PAGE

IV. ASCERTAINING DIMENSIONALITY 56

Structure Statistics in Review 58
A Real Live Example 59
Artificial Structure Approximations of the Actual Data 60

V. SUMMARY 63

MDS and Educational Measurement 63
Objectives 64
Procedure 65
The Structure Statistic 65
Final Conclusions 66

REFERENCES 68

LIST OF FIGURES

FIGURE		PAGE
1	Plots of percentage of total variance accounted for by each of the first 17 factors for ninth and twelfth grade boys	11
2	Mean root sizes for 100 15 x 35 random input MDS runs	16
3	Distribution of Alpha statistics for random input runs	18
4	Alpha as a function of the additive constant for five runs with 15 objects and 35 judges	20
5	Distribution of Beta statistics for random input runs	22
6	Beta as a function of the additive constant for five runs with 15 objects and 35 judges	23
7	Distribution of Gamma statistics for random input runs	25
8	Gamma as a function of the additive constant estimate for five sets of random data	27
9	Distribution of Delta statistics for random input runs	28
10	Delta as a function of the additive constant	30
11	Distribution of Epsilon statistics for random input runs	32
12	The Algebraic sum (Epsilon) and the absolute sum of roots as a function of different estimates of the additive constant	33
13	Size of characteristic roots with various estimates of the additive constant for one set of MDS data	35
14	Alpha as a function of average interpoint distance with one nonerror dimension	40
15	Beta as a function of average interpoint distance with one nonerror dimension	41
16	Gamma as a function of average interpoint distance with one nonerror dimension	42
17	Delta as a function of average interpoint distance with one nonerror dimension	43

FIGURE		PAGE
18	Epsilon as a function of average interpoint distance with one nonerror dimension	44
19	A graphic comparison of five structure statistics as one dimensional structure is built up	45
20	Structure statistics for the largest root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in	48
21	Structure statistics for the smallest nonerror root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in.	51
22	Mean root sizes for 15 x 35 MDS runs with different amounts of nonerror structure built in.	53
23	Structure statistics for the largest error root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in	54
24	The components of nonerror variance and error variance alone and in combination.	57
25	Approximations to actual-data MDS Run X by runs with artificially built-in structure.	61

LIST OF TABLES

TABLE		PAGE
1	The Sensitivity of Five Statistics as Revealed by the Distance Each is Moved into the Critical Region by Building-in Unidimensional Nonerror Structure	47
2	Root Sizes with Varying Amounts of Unidimensional Nonerror Structure	52
3	Structure Statistics for Actual MDS Problem	60

CHAPTER I

MULTIDIMENSIONAL SCALING

During recent years, psychologists and educators have become increasingly aware of the multidimensional nature of many of the phenomena in their fields. A number of models have been proposed and methodologies developed to help the research worker describe multidimensional phenomena.

One of the models that is currently being utilized is the multidimensional scaling model derived by Richardson (1938) with the aid of Young and Householder (1938) and formalized by Torgerson (1952; 1958). This model involves the location of stimulus objects in a "psychological space" and the factor analytic identification of the dimensions of the space. The stimulus objects can be things, persons, groups or events. Scaling of these objects is merely an attempt to describe them.

The investigator who chooses to utilize Torgerson's multidimensional scaling (MDS) model finds that there are some phases of the technique that lack refinement. Two problems seem to be especially bothersome.

The first problem occurs in the factor analytic determination of the dimensionality of the psychological space. Factor analysis is based upon certain mathematical assumptions. Unfortunately, such assumptions often do not coincide with circumstances in the real world. This is the case with factor analysis. A factor analysis of infallible data, i.e., error-free measurements, will readily identify the dimensionality of the space defined by a set of points; however, when the location of the points in the space is erroneously determined, i.e., when the measurements contain error, the factor analysis will identify one less dimension than there are points. Consider a set of n points, for which a factor analysis has identified (n-1) dimensions, where only k dimensions can be considered

nonerror dimensions. For this factor analysis, (n-1-k) dimensions may be considered artifacts of the measurement error or error dimensions. A procedure for determining k, the number of nonerror dimensions, is needed.

A second problem occurs in carrying out the MDS computation. The MDS model involves a set of points in a perceptual or judgmental space. Each point represents a stimulus object. The interpoint distances represent psychological distances between the stimuli. The computed interstimulus distances, however, are relative rather than absolute expressions of distance. Relative interstimulus distances may be converted to absolute interpoint distances by the addition of a constant to the relative distances. Messick and Abelson (1956) indicate that the addition of a constant to a set of relative distances will not affect the validity of the psychological space, i.e., the configuration. However, adding the constant will affect the amount of variance associated with every root, and so the results may be distorted if an erroneous constant is added. At the present time there is no direct solution for determining the appropriate size of the additive constant. The additive constant must be estimated. The effects of errors in estimating this constant are not fully known, but the very fact that an estimate is necessary makes this MDS procedure suspect.

Multidimensional analysis may eventually be made by quite different procedures than those outlined by Torgerson. Shepard (1961) has developed one alternative. Because it was believed that ultimate methodological preferences should be based on empirical findings as well as theoretical rationales, the present study was undertaken.

Objectives of This Study. The primary purpose of this study was to consider the first of the two problems cited above, that of determining the number of nonerror dimensions. The answer to the problem lies in the

knowledge of how error variance is distributed in the multidimensional space. Though the error is random, the dimensions will surely not be loaded with equal amounts of error.

This problem is not unlike the traditional problem in statistical inference, "Is this outcome attributable to chance or is it very unlikely that this outcome could occur with random input data?" The answer is traditionally provided by exposition of the outcomes due to chance alone, an answer that mathematical statisticians can provide. As a result of their study of probability, the distributions of chi square, t, F, and other statistics have been generated; and the traditional precise but hedged answer to the above question has been given many, many times.

To solve the MDS problem we need similar, theoretically-determined, sampling distributions. However, the MDS procedure involves twisting and turning of the data, and few mathematical statisticians are attracted by such inelegant problems. It appeared that any relevant sampling distributions would have to be obtained empirically; by scaling random "judgments" again and again. Exactly that was done in this study.

A search was undertaken to find one or more descriptive statistics to aid in the discrimination between nonerror dimensions and error dimensions. In succeeding chapters these statistics are discussed with regard to the structure of random data, artificial structure, and structure from real data. Before the details of the procedure that was followed are spelled out, the general rationale of Torgerson's MDS approach will be discussed.

Theoretical Background. Traditional psychological scaling procedures are based upon knowledge of, or assumptions about, the dimensionality of the attribute being scaled. When these procedures are used, the subject is asked

to indicate which of two stimuli has more of the given attribute. To utilize one of these procedures it is necessary for the experimenter to define the attribute. For meaningful results it is necessary for the subjects to understand the experimenter's definitions. Only then can there be assurance that the subjects' judgments of "inequality" are applicable to a psychological domain.

When dealing with psychological characteristics these questions are raised, "Can the experimenter adequately define the attribute?" and "Does the experimenter specify the attributes that the subject considers most relevant?" Certainly not always. In the measurement of attitudes, for example, the commonly-used, unidimensional procedures require the above mentioned attribute definition. Messick (1956c) provides evidence, however, that not all individuals consider only a single dimension in arriving at the required decisions. He found that many persons consider two dimensions. Some no doubt consider more. Forcing subjects to respond to a single, experimenter-defined characteristic may not only obscure relevant information but may distort the obtained information.

Multidimensional scaling procedures purportedly alleviate, if not eliminate, some of the disadvantages of unidimensional scaling. But when the multidimensional techniques are utilized, various additional complexities must be dealt with. It is felt, however, that at least for some purposes the advantages of MDS outweigh the disadvantages (Messick, 1956b).

Multidimensional scaling utilizes a psychological space that is assumed to have Euclidean characteristics. There has been some debate as to whether or not the Euclidean model is an appropriate one for representing psychological processes; however, there is little evidence for rejecting the Euclidean model and much reason to support its use. (c. f. Helm, 1959; Messick, 1956b; Morton, 1959; Richardson, 1938; Torgerson, 1952).

The psychological space is defined in terms of interstimulus distances. These distances are determined on the basis of judgments of similarity between stimuli, i. e., the psychological distance (d_{ij}) between stimuli $\underline{i}$ and $\underline{j}$ is an inverse function of the similarity seen in the two stimuli. Gulliksen (1961) points out that subjects can make such judgments with a relative degree of accuracy and stability and that the experimenter is relieved of the necessity of defining the attribute being considered.

Young and Householder's Solution

Prior to the work of Young and Householder (1938), a set of points was customarily described in Cartesian terms, i. e., according to the coordinates in the space or their projections on the reference axes of that space. Thus, each point was located by what amounted to a scale value for each dimension, e. g., if there were two dimensions point $\underline{j}$ would be located as $(X, Y)_j$; point $\underline{k}$ as $(X, Y)_k$, etc., where X and Y are projections of a point on two axes. Young and Householder presented a technique for describing the location of points in the space in terms of interpoint distances and then presented a proof that the solution for the dimensionality of that space was equivalent to the solution for the dimensionality of a space in which the points were located by their Cartesian coordinates. Basic to the proof was the expression of interpoint distance in terms of Cartesian coordinates.

The space was assumed to be Euclidean. In the Euclidean spatial model, the distance (d_{jk}) between points $\underline{j}$ and $\underline{k}$ is equal to the square root of the sum of the squared distances in each plane. In a space of $\underline{r}$ dimensions, this distance model may be expressed by the following equation:

$$d_{jk}^2 = \sum_{i=1}^{r} (a_i - b_i)^2, \qquad (1)$$

where $\underline{a}$ and $\underline{b}$ are the projections of points $\underline{j}$ and $\underline{k}$ respectively on the various axes, $\underline{i}$, i. e., the Cartesian coordinates of points $\underline{j}$ and $\underline{k}$.

Young and Householder converted the matrix of interpoint distances to a matrix of scalar products, B, whose elements (b_{ij}) are defined by the following equation:

$$b_{ij} = \tfrac{1}{2}(d_{in}^2 + d_{jn}^2 - d_{ij}^2) \qquad (2)$$

where point $\underline{n}$ is any point in the space and is used as a reference point. Their proof continues by showing that the dimensionality of the space is equal to the rank of the B matrix. There will be a unique matrix, B, for each point selected as $\underline{n}$; however, with error-free data the dimensionality remains unchanged with various B matrices.

The matrix of scalar products, B, was shown by Young and Householder to be positive and semi-definite, which will insure that the set of points has a unique dimensionality in a real Euclidean space. This is true, of course, only for data that contain no error. The presence of error generally precludes the possibility of obtaining a B matrix that is positive semi-definite, thereby precluding the possibility of obtaining a solution with real dimensionality in the Euclidean sense. In other words, the conditions demanded by the Young and Householder solution cannot be met when data contain error. Since the data obtained from most experimental sources contain some error, the researcher must make some assumptions about the data and about the final statement of dimensionality if he is to utilize the Young and Householder solution. The assumption made by Torgerson (1952) was that with fallible data, only the large latent roots (characteristic roots) represent the true dimensionality of the space and that the smaller roots are artifacts of the error in the data. Such an assumption has not been difficult to make; however, we are not prepared to specify what constitutes a large root.

Torgerson's B* Matrix

When fallible data are used, each point is located in the space with some error. Thus, a different matrix, B, may be obtained for the different points which may be designated as point n in equation (2). Torgerson (1952) reasoned that if the centroid of the points in the space was used rather than some point n, the errors would nullify each other and a more satisfactory solution for the matrix of scalar products would result. The B* matrix is a matrix of scalar products with the reference at the centroid of points in the space.

The B* matrix was the focus of analysis in the study reported herein.

Torgerson's MDS Solution

Although Young and Householder (1938) presented the methodology for determining the dimensionality of a set of points in terms of their interpoint distances, and Richardson developed the model for the determination of the interpoint distances (Richardson, 1938), apparently only Klingberg utilized the technique (Klingberg, 1941) before 1952 when Torgerson explicitly stated the procedures for MDS (Torgerson, 1952). Torgerson presented MDS as involving three basic steps. "In the first step, a scale of comparative distances between all pairs of stimuli is obtained. The second step involves estimating an additive constant and using this estimate to convert the comparative distances into absolute distances. In the third step, the dimensionality of the psychological space necessary to account for these absolute distances is determined, and the projections of the stimuli on axes of this space are obtained." (Torgerson, 1952).

The first step can be adequately handled by the method of "pair-comparison" or an extension thereof, e. g., the method of triadic combinations, the complete method of triads, etc. Torgerson suggested the complete method

of triads. Klingberg (1941) found that the multidimensional method of rank-order was more satisfactory in an experimental situation in which the other methods mentioned above were also used.

The first step involves the use of Thurstone's Law of Comparative Judgment (1927) or some such transformation of frequencies of judgments into psychological distances. Hefner (1958) has opposed the use of the Gaussian distribution, supporting instead a non-central chi square distribution. Most of the field work has followed Thurstone's use of the Gaussian curve.

The second step, that of determining an additive constant that will convert relative interpoint distances to absolute distances in a Euclidean space where the number of dimensions is a minimum, has been more difficult to overcome. If the B* matrix is known, the value of the additive constant, c, that yields, the largest number of zero roots in the characteristic equation,

$$\left| B^* - RI \right| = 0 \qquad (3)$$

(Thurstone, 1947) with the remaining roots all positive, is the one desired. However, c is necessary for the construction of B*. Thus, a direct solution is not possible and c must be estimated.

An iterative procedure which would probably result in a quite accurate estimate of c has been provided by Messick and Abelson (1956). However, this is applicable only when it is certain there is some nonerror variance. Torgerson (1952; 1958) recommended the way of estimating c that was used in the present study. His rationale is presented here. For criticism of Torgerson's method, see Messick and Abelson (1956) and Hefner (1958).

Estimating the Additive Constant. Torgerson's initial assumption necessary for the estimation of the additive constant, c, was that in a space of limited dimensionality and containing many points, four points could be found that aligned

themselves along a single dimension. The additive constant that would be appropriate for this unidimensional sub-space, since the sub-space was part of the total space, would be a good estimate of the desired additive constant. For any three points whose relative interpoint distances are known, a constant may be determined that will place the points in a straight line such that

$$d_{ik} = d_{ij} + d_{jk}. \tag{4}$$

Equation (4) is a necessary condition for distance in a Euclidean space. If we designate the relative interpoint distance from i to j as h_{ij}, and the constant that will place points i, j, and k in a straight line as c_{ijk}, the equation for the determination of c_{ijk} is

$$c_{ijk} = h_{ik} - h_{ij} - h_{jk} \tag{5}$$

(Torgerson, 1958). For a set of n stimuli n(n-1)(n-2)/6 values for c_{ijk} may be determined; one for each different set of three stimuli. Torgerson (1952) noted (a) that generally points most nearly in a straight line will give the largest value for c_{ijk}; (b) if four points, i, j, k, and l, can be found such that c_{ijk}, c_{ijl}, c_{ikl}, and c_{jkl} give the highest estimates of c in a consistent manner, the four points can be considered to be lying in a unidimensional sub-space; and (c) even if such a set of four points cannot be found, the largest value of c_{ikl} might still be a worthwhile estimate of c. To obtain a slightly more conservative estimate, the fifth largest value for c_{ikl} was used as the estimate of c in this study.

Factor Analysis. Torgerson (1952) identified the third step in the solution of the multidimensional scaling problem as being the identification of the dimensionality of the space that is necessary to account for the absolute interpoint distances and the determination of the projections of the points on the axes of the space. This may be accomplished by any one of the factor analytic procedures.

Solving for the characteristic roots and vectors of the B* matrix is perhaps the most direct factor solution. (The characteristic roots are sometimes referred to as Eigen roots or latent roots). In matrix symbols,

$$B^* = VRV' \tag{6}$$

where V is the matrix of vectors and R is the diagonal matrix of roots. It is apparent that the principal axes factor matrix, F, can be obtained from these roots and vectors by

$$F = VR^{\frac{1}{2}} . \tag{7}$$

Where the dimensionality of the B* matrix is the prime issue the roots matrix serves as well as the factor matrix. In this study the distribution of error variance will be investigated by examination of the characteristic roots and various statistics derived from the characteristic roots.

Each characteristic root can be considered a component of variance of the distances between stimulus objects. The largest root then represents the variance associated with the maximum dimension, that mathematically unique dimension along which more variability in objects occurs than any other possible dimension. The second root represents the maximum dimension in the space orthogonal to the first dimension, and so on.

Sjogren (1961) used the characteristic roots to study the organization of mental abilities of students enrolled in secondary schools. Some of his data are presented in Figure 1. It is felt that these results are typical of those one might obtain from an MDS study. The plots in Figure 1 represent the percentage of variance accounted for by the largest 17 roots for two groups of subjects. The curve for the twelfth grade boys is very regular after the 6th root. According to the subjective methods now employed, an attempt would probably be made to interpret the largest 6 roots. A conservative investigator might see justification

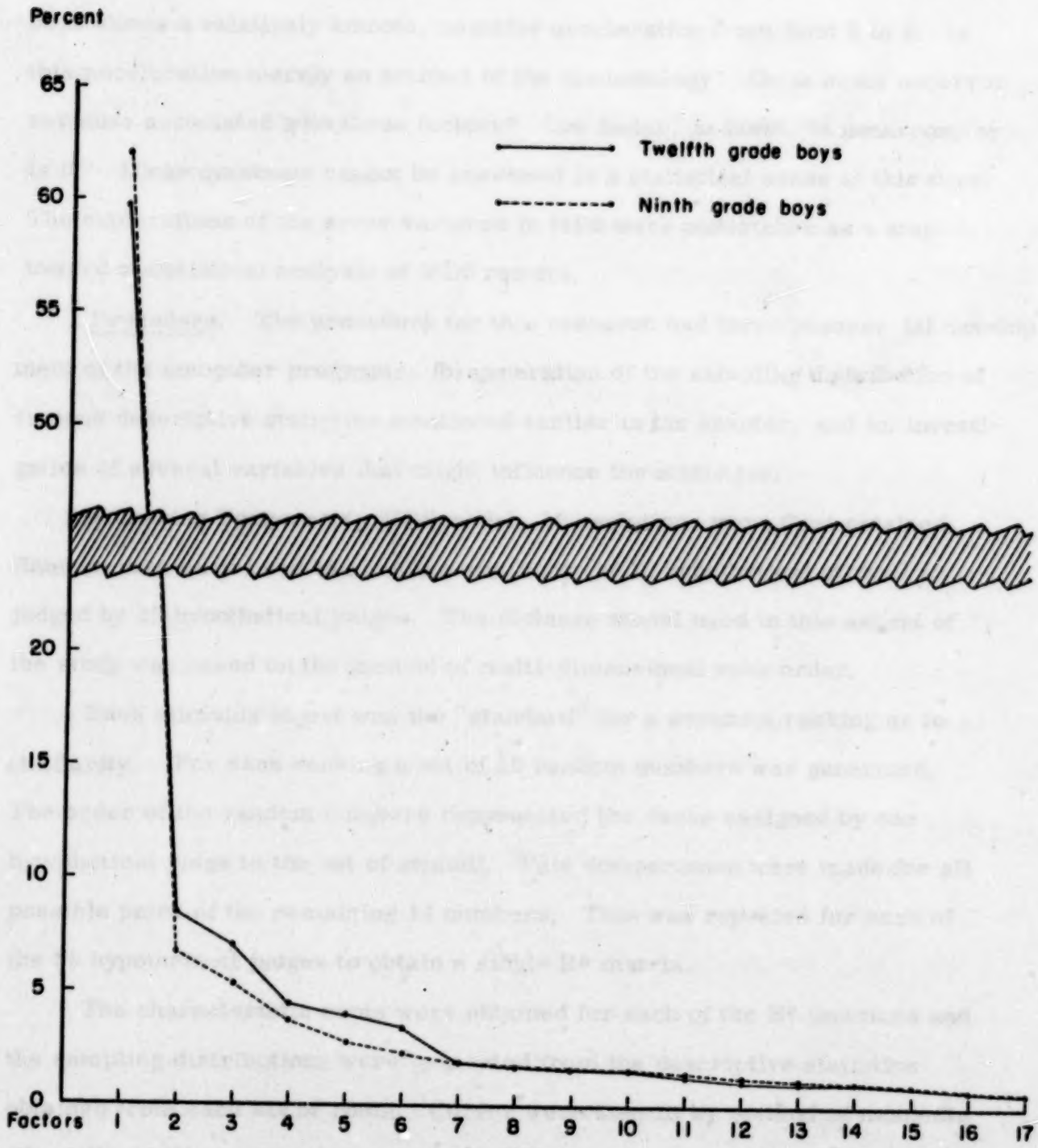


Fig. 1. Plots of percentage of total variance accounted for by each of the first 17 factors for ninth and twelfth grade boys. (Sjogren, 1961)

for attempting to interpret only Roots 1, 2, and 3. The curve for ninth grade boys shows a relatively smooth, negative acceleration from Root 2 to 6. Is this acceleration merely an artifact of the methodology? Or is some nonerror variance associated with these factors? One factor, at least, is nonerror, or is it? These questions cannot be answered in a statistical sense at this time. The explorations of the error variance in MDS were undertaken as a step toward a statistical analysis of MDS results.

Procedure. The procedure for this research had three phases: (a) development of the computer programs, (b) generation of the sampling distribution of various descriptive statistics mentioned earlier in the chapter, and (c) investigation of several variables that might influence the statistics.

Following Torgerson's MDS model, 100 solutions were first obtained. Random data were used to simulate similarity of 15 hypothetical stimuli as judged by 35 hypothetical judges. The distance model used in this aspect of the study was based on the method of multi-dimensional rank order.

Each stimulus object was the "standard" for a separate ranking as to similarity. For each ranking a set of 15 random numbers was generated. The order of the random numbers represented the ranks assigned by one hypothetical judge to the set of stimuli. Pair comparisons were made for all possible pairs of the remaining 14 numbers. This was repeated for each of the 35 hypothetical judges to obtain a single B* matrix.

The characteristic roots were obtained for each of the B* matrices and the sampling distributions were generated from the descriptive statistics obtained from each set of roots. Curves were then fit by method of moments (Kendall, 1942) to these distributions and the critical points for a one-tailed test of significance were obtained from appropriate tables (Pearson, K., 1934, 1951; Pearson, E. S., 1922; Brownlee, 1923; Pairman, 1919).

Phase II. In the next phase of the study the influence of certain variables on the statistics was investigated. Included among the variables investigated were (a) the number of stimulus objects judged, (b) the number of judges, (c) the size of the additive constant, and (d) the number and size of artificial dimensions built into the solutions.

To investigate the effect of the size of the additive constant on the statistics, five cases were drawn from the sample of 100. Twelve estimates of the additive constant, ranging from .0000 to 2.0000 were used to construct the B* matrix. Thus twelve B* matrices were produced for each of the five cases. The characteristic roots of these B* matrices were obtained and the results examined.

Nearly 100 additional solutions were then obtained. In some of these solutions the number of stimulus objects involved was varied or the number of judges was varied to determine whether this affected the descriptive statistics. Also, artificial dimensionality varying from one to three dimensions was built into many of the solutions for the same purpose. The method used in the last mentioned solutions was essentially the same as for the first 100 solutions except that random normal deviate values were used as the hypothetical stimuli rather than random numbers. The deviate values were used to provide a meaningful scale by which artificial dimensionality could be built into the solution.

In addition to the above mentioned solutions, two solutions were computed using the complete method of triads. This was accomplished by generating three random numbers as hypothetical stimuli for every triadic judgment made by each judge. Two solutions based on actual data were also investigated.

More detail on procedure will be presented in the following chapters as the results of the different phases are reported.

CHAPTER II

STRUCTURE STATISTICS

For each analysis of data of an MDS problem the results may be characterized in terms of certain descriptive statistics. Most of the possible statistics would be characteristic roots or derivatives of the characteristic roots and vectors of the B* matrix. Since these statistics will be used in the description of the structure of the MDS space, they will be called structure statistics. At the outset it was not known which of the structure statistics would be useful.

A general solution for characteristic roots. The processing of characteristic roots as described in the following sections of this chapter would not have been necessary if the B* matrices from random data had met certain specifications. Various properties of characteristic roots and vectors have been studied for matrices whose entries can be considered random selections from a normal distribution (Mallows, 1961). With B* matrices meeting those specifications, the focus of this study would have been on the properties of the empirically obtained B* matrices rather than on the properties of the empirically obtained characteristic roots and vectors as described in terms of the structure statistics.

Frequency distributions of specifically located entries from 100 B* matrices were plotted. It was found that the corresponding cells of different matrices were distributed approximately normally, but that the off-diagonal cells were distributed around zero and the diagonal cells, all positive, of course, were distributed around a point about three standard deviation distances above zero. The entries did appear to be random (uncorrelated) but because of the discrepancy between diagonal and off-diagonal entries the Mallows analysis was not deemed relevant.

Multivariate characteristic roots and vectors can sometimes be analyzed in terms of Wishart distributions (Anderson, 1958). The authors were advised,

however, that the B* matrices were unlike those for which the Wishart generalized t^2 distributions were appropriate.

It was believed that existing mathematical studies of characteristic roots and vectors would not provide a basis for defining and evaluating MDS structure statistics.

Empirical distributions. In order to evaluate the structure statistics when the structure is based entirely on error variance, the Torgerson solution for MDS was applied again and again to random numbers. These random numbers were treated as if they were judgments of similarity of objects as perceived by judges. The special case of 15 stimulus objects and 35 judges was used for the first 100 MDS solutions to provide sampling distributions. The mean root sizes for these 100 runs are shown in Figure 2. A line joining the tops of the bars would be almost rectilinear.

Five different structure statistics were calculated for each solution. They will be explained and evaluated in this chapter. To be most useful such a statistic should be accessible, unbiased, efficient, sensitive to fluctuations in nonerror variance, and relatively unaffected by errors in estimating the additive constant. Some of the five statistics are less accessible than others in that they require computation of more characteristic roots. Bias cannot be determined since population parameters are unknown. Efficiency is ordinarily indicated by the variance of the statistic, the greater the variability the less the efficiency. Because each different statistic here would be used to estimate a different parameter, no direct comparison as to efficiency was possible. Sensitivity was determined by observing how much structure had to be "built-in" to yield a statistically significant value. This consideration will be reported in the next chapter. A desirable but unessential characteristic of a structure statistic

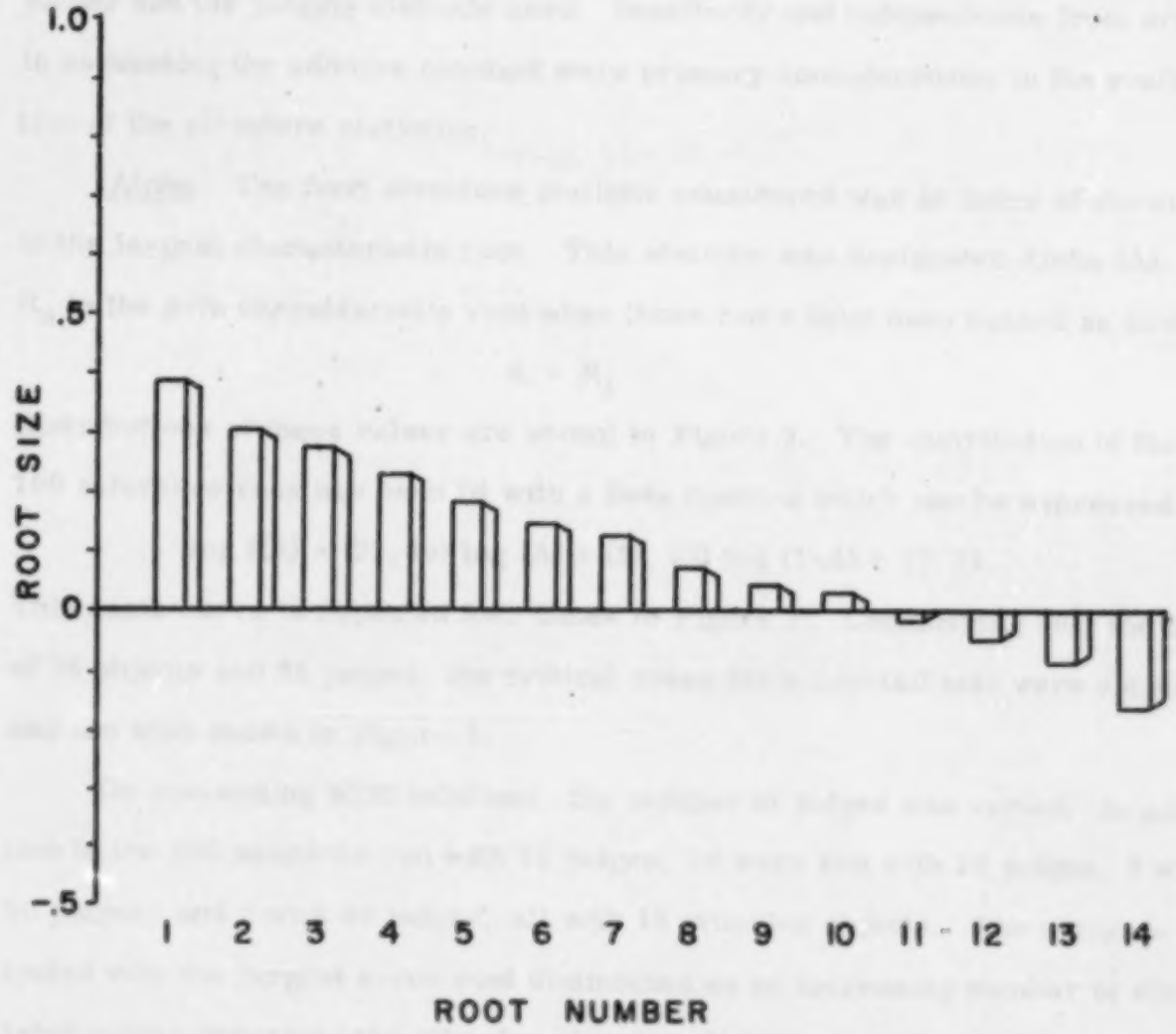


Fig. 2. Mean root sizes for 100 15 x 35 random input MDS runs.

would be an independence from the number of stimulus objects, the number of judges and the judging methods used. Sensitivity and independence from errors in estimating the additive constant were primary considerations in the evaluation of the structure statistics.

Alpha. The first structure statistic considered was an index of the size of the largest characteristic root. This statistic was designated Alpha (A). If R_n is the n-th characteristic root when those roots have been ranked as to size,

$$A = R_1 \quad . \tag{8}$$

Distributions of these values are shown in Figure 3. The distribution of the 100 reference runs has been fit with a Beta function which can be expressed as

$$\log f(A) = (22.49) \log (A) + (36.13) \log (1-A) + 17.71. \tag{9}$$

This same curve is repeated four times in Figure 3. Considering only the case of 15 objects and 35 judges, the critical areas for a one-tail test were estimated and are also shown in Figure 3.

On succeeding MDS solutions, the number of judges was varied. In addition to the 100 solutions run with 35 judges, 10 were run with 20 judges, 3 with 50 judges, and 2 with 80 judges, all with 15 stimulus objects. The variance associated with the largest error root diminished as an increasing number of simulated judges perceived the stimulus objects. If Alpha is to be used as a structure statistic, the number of judges appears to be an important variable.

The number of stimulus objects was reduced from 15 to 8 and, with 35 judges, an additional ten runs were completed. The results of this variation are shown in Figure 3. When the random-input B^* Matrix is reduced from 15 x 15 to 8 x 8 there is less total variance, and consequently, less variance associated with the largest root. The mean Alpha dropped from .39 to .19 as shown in the third plot in Figure 3.

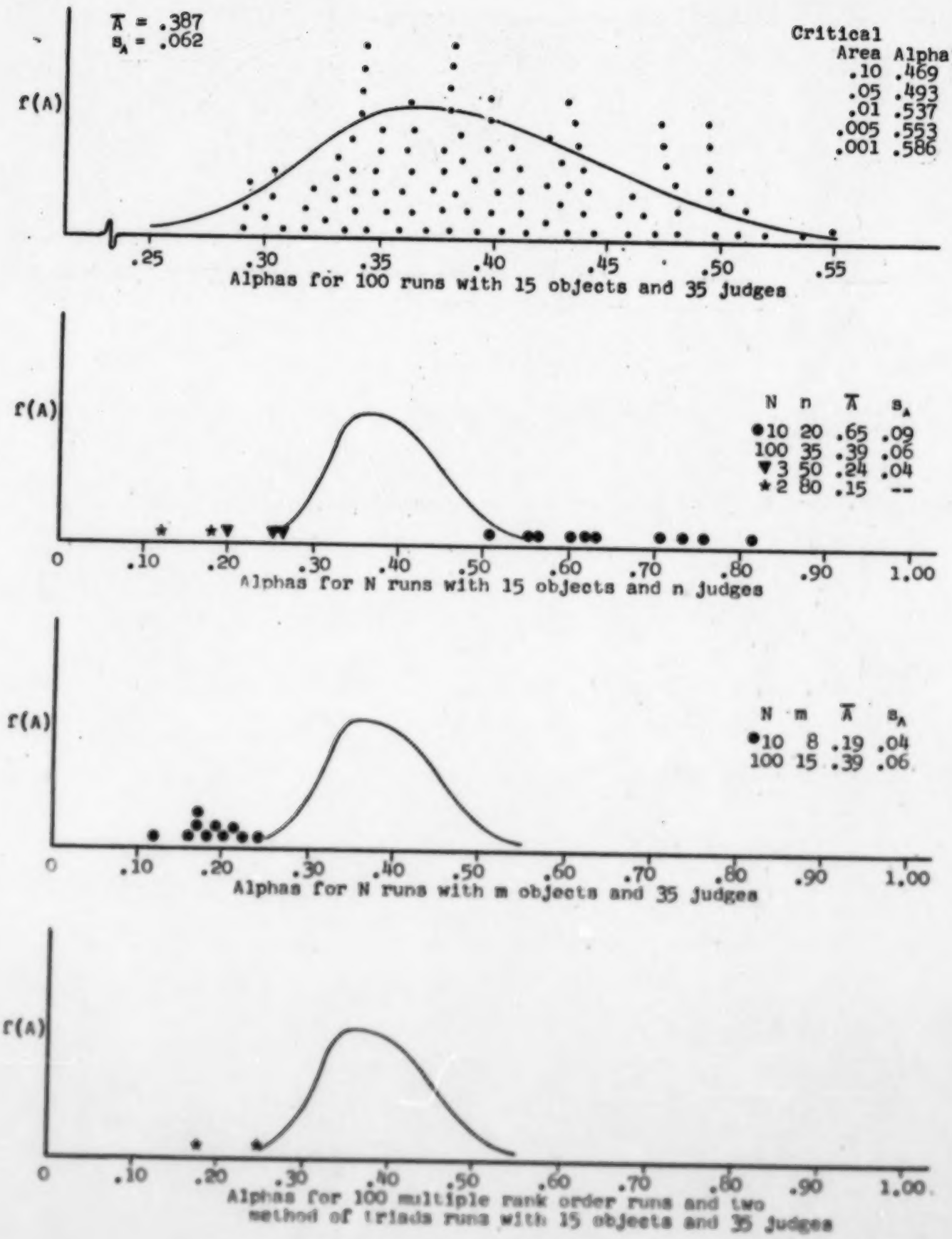

$\bar{A}$ = .387
s_A = .062
f(A)
Critical
Area Alpha
.10 .469
.05 .493
.01 .537
.005 .553
.001 .586
.25 .30 .35 .40 .45 .50 .55
Alphas for 100 runs with 15 objects and 35 judges
f(A)
N n $\bar{A}$ s_A
●10 20 .65 .09
100 35 .39 .06
▼3 50 .24 .04
★2 80 .15 --
0 .10 .20 .30 .40 .50 .60 .70 .80 .90 1.00
Alphas for N runs with 15 objects and n judges
f(A)
N m $\bar{A}$ s_A
●10 8 .19 .04
100 15 .39 .06
0 .10 .20 .30 .40 .50 .60 .70 .80 .90 1.00
Alphas for N runs with m objects and 35 judges
f(A)
0 .10 .20 .30 .40 .50 .60 .70 .80 .90 1.00
Alphas for 100 multiple rank order runs and two
method of triads runs with 15 objects and 35 judges

Fig. 3. Distribution of Alpha statistics for random input runs.
--Each of the four curves above represents the same 100 reference runs.

When the method of judging was changed from multiple rank order to the method of triads, the 35 simulated judges found the 15 simulated objects to occupy a much reduced psychophysical space. This reduced space is evidenced in part by the two relatively small Alphas shown at the bottom of Figure 3.

In order to determine the influence of errors in the additive constant on Alpha, five of the runs were selected for systematic variation of the additive constant. Although additive constant estimates would almost certainly be between .4 and .6 for these solutions, estimates ranging from 0 to 2.0 were considered. The results are displayed in Figure 4. When best estimates are used, about 94% of the Alphas would be expected to lie in the shaded zone. It appears that if there is an overestimation of .1, a quite unlikely amount, Alphas which should have been at the .05 level of significance will appear about at the .001 level of significance. Even with smaller mistakes there will be distortion. Overestimates will lead to Type I errors and underestimates to Type II errors in evaluating MDS solutions with Alpha.

To summarize the information of the preceding paragraphs, the Alpha statistic is accessible, but obviously sensitive to changes in the number of stimulus objects, changes in the number of judges, changes in the method of judging, and to errors in estimation of the additive constant.

Beta. A ratio of largest root size to second largest was selected as the second structure statistic. This ratio was designated Beta, (B).

$$B = R_1/R_2 \ . \qquad (10)$$

Beta cannot be less than unity. It will tend to be large when R_1 is a nonerror root and R_2 is an error root.

One hundred Betas were obtained at the same time the Alphas were obtained. The distribution of these Betas are shown in Figure 5. Each dot in the upper

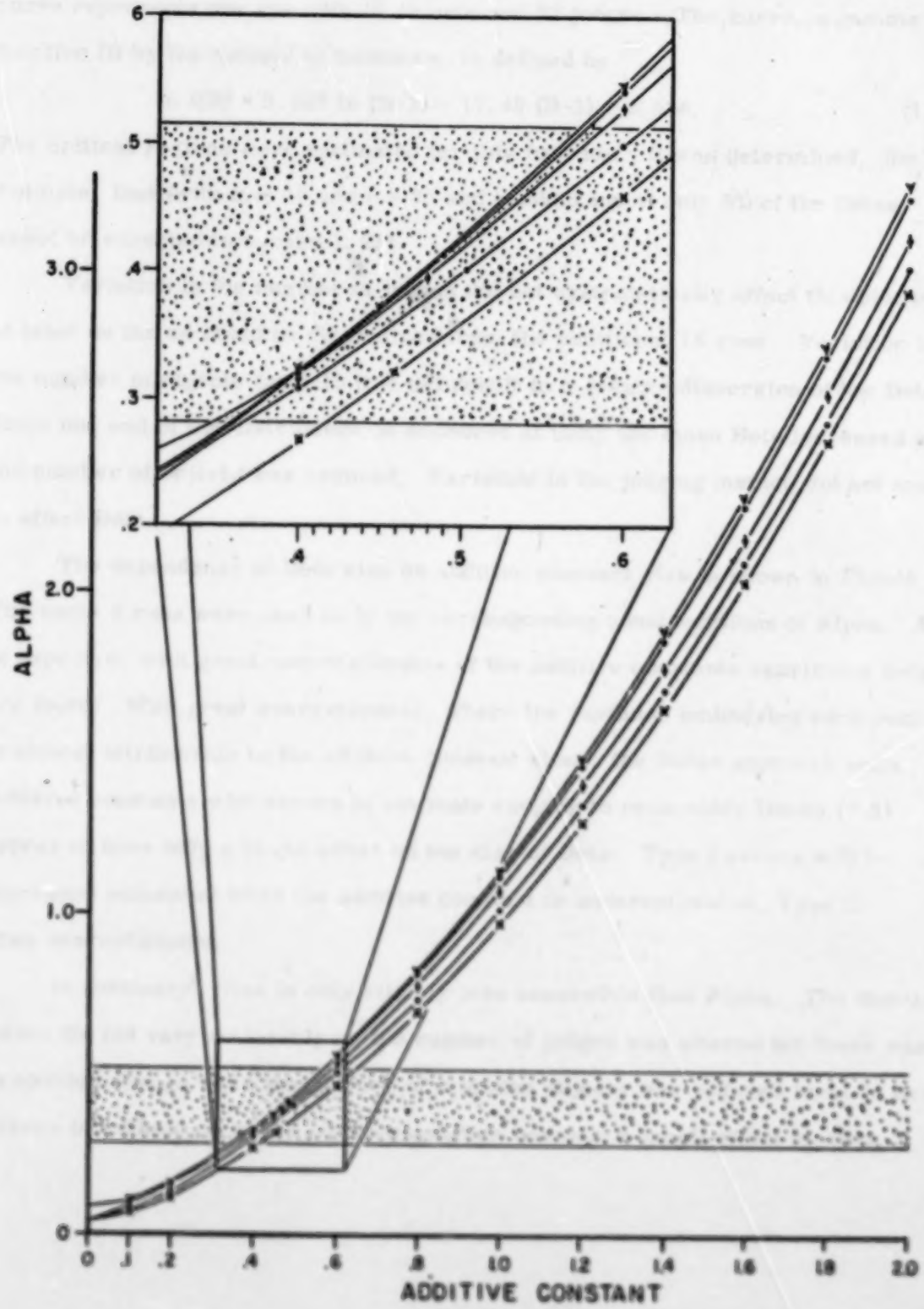


Figure 15

Fig. 4. Alpha as a function of the additive constant for five runs with 15 objects and 35 judges.

curve represents one run with 15 objects and 35 judges. The curve, a gamma function fit by the method of moments, is defined by

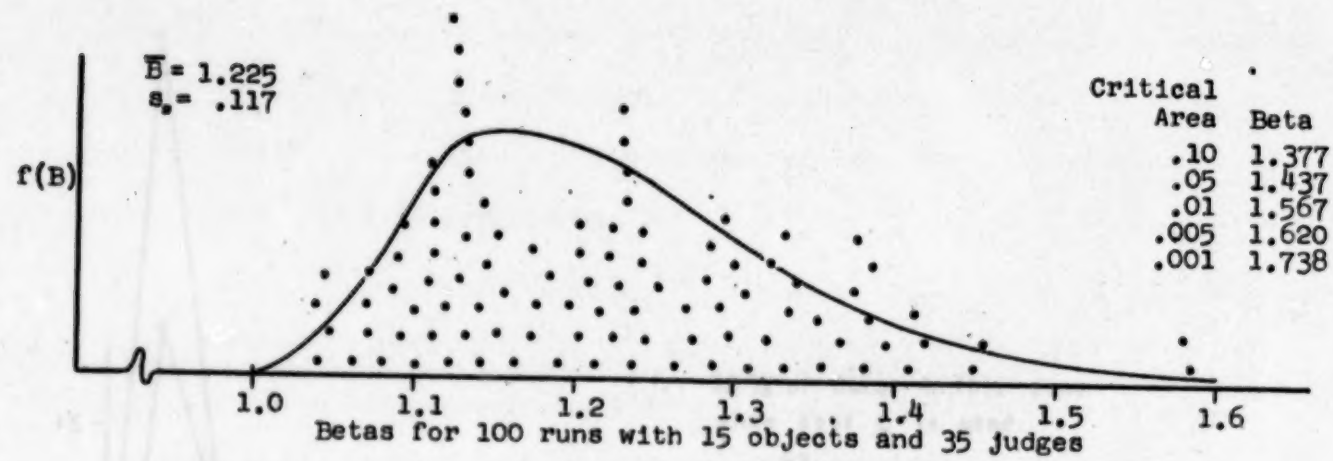
$$\ln f(B) = 2.937 \ln (B-1) - 17.49 (B-1) + 9.548. \quad (11)$$

The critical regions were estimated for this function. It was determined, for example, that (with m = 15 and n = 35 and random input) only 5% of the Betas would be expected to exceed 1.437.

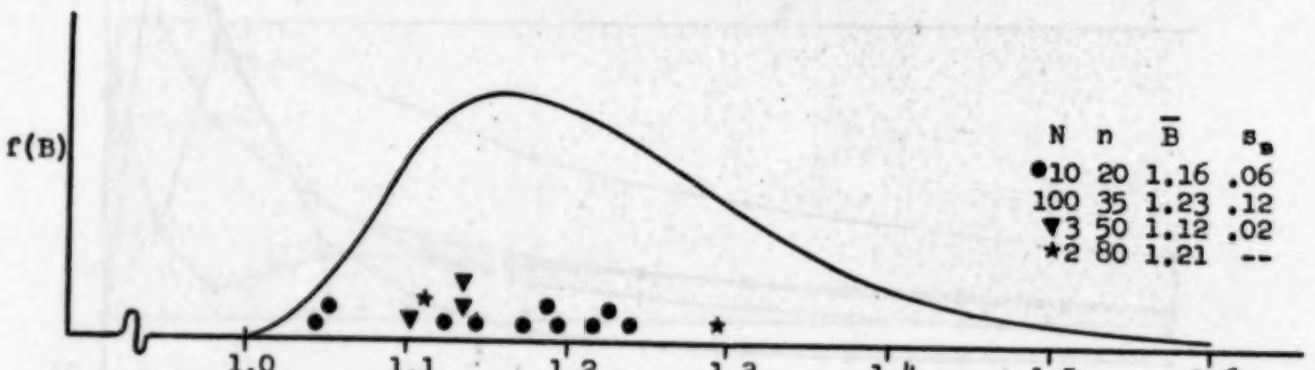
Variation in the number of judges did not systematically affect this statistic, at least as far as could be determined from the additional 15 runs. Variation in the number of objects from 15 to 8 did result in a greater dispersion of the Betas. Since one end of the distribution is anchored at unity the mean Beta increased as the number of objects was reduced. Variation in the judging method did not appear to affect Beta.

The dependence of Beta size on additive constant size is shown in Figure 6. The same 5 runs were used as in the corresponding considerations of Alpha. As is expected, with great underestimates of the additive constants capricious Betas, are found. With great overestimates, where the variance embodying each root is almost attributable to the additive constant alone, the Betas approach unity. Additive constants with errors of estimate ranging to reasonable limits ($\pm .1$) appear to have only a slight effect on the size of Beta. Type I errors will be increased somewhat when the additive constant is underestimated, Type II when overestimated.

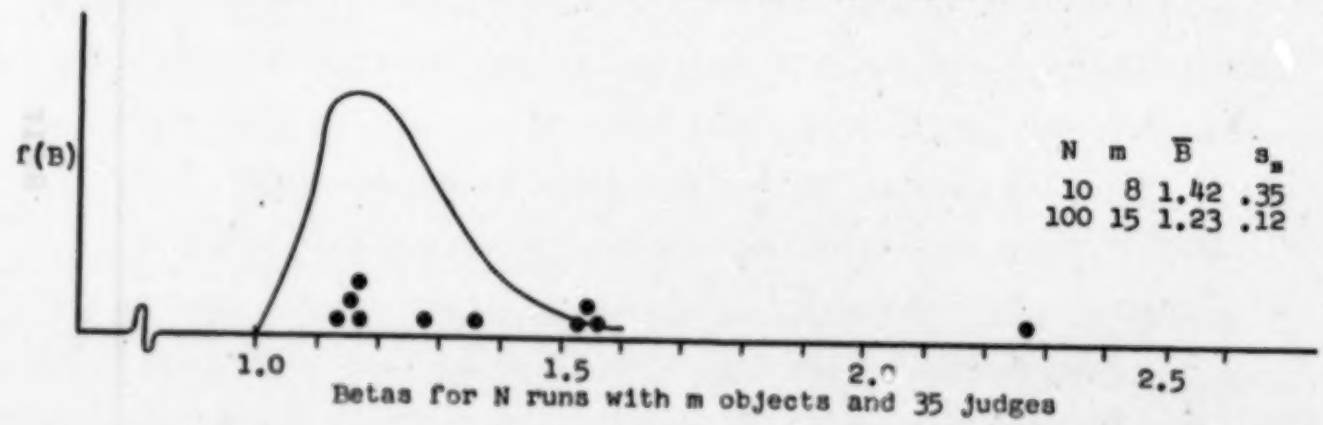
In summary, Beta is only slightly less accessible than Alpha. The distribution did not vary noticeably as the number of judges was altered but there was an obvious change when the number of stimulus objects was reduced from 15 to 8. Errors in estimation of the additive constant had only a small effect on Beta.



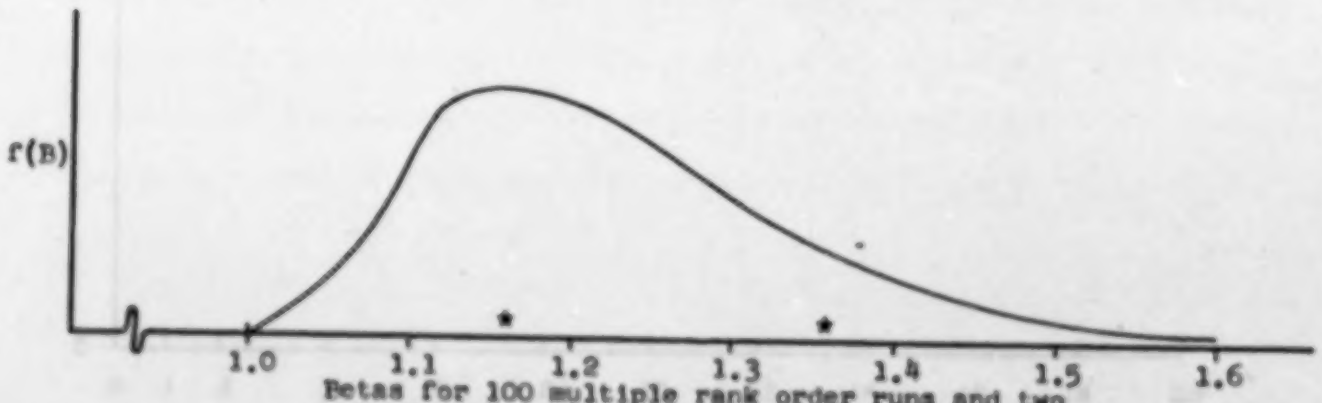
Betas for 100 runs with 15 objects and 35 judges



Betas for N runs with 15 objects and n judges



Betas for N runs with m objects and 35 judges



Betas for 100 multiple rank order runs and two method of triads runs with 15 objects and 35 judges

Fig. 5. Distribution of Beta statistics for random input runs.
--Each of the four curves above represents the same 100 reference runs.

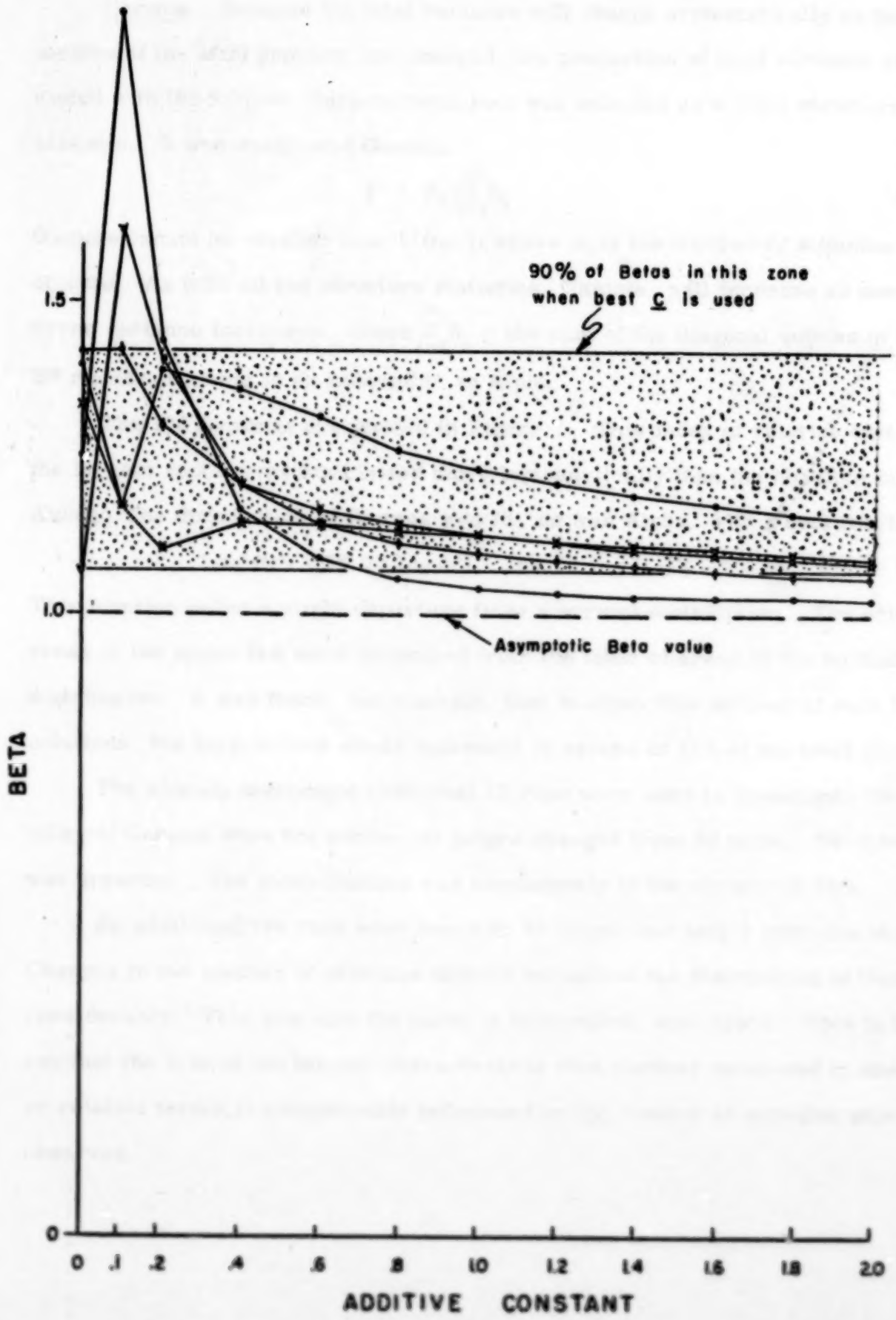


Fig. 6. Beta as a function of the additive constant for five runs with 15 objects and 35 judges.

Gamma. Because the total variance will change systematically as parameters of the MDS problem are changed, the proportion of total variance associated with the largest characteristic root was selected as a third structure statistic. It was designated Gamma.

$$\Gamma = R_1 / \sum_{i=1}^{m} R_i \qquad (12)$$

Gamma cannot be smaller than 1/(m-1) where m is the number of stimulus objects. As with all the structure statistics, Gamma will increase as non-error variance increases. Since $\sum_{i=1}^{m} R_i$ = the sum of the diagonal entries of the B* matrix, Gamma is as accessible as Alpha.

The 100 Gammas are plotted in Figure 7. This index of relative size of the largest root seems distributed more leptokurtically than the absolute index, Alpha. The distribution of Gamma was fit, as was Alpha, with a beta function.

$$\text{Log } f(\Gamma) = 56.22 \log (\Gamma) + 164.72 \log (1-\Gamma) - 55.74. \qquad (13)$$

This function is but a slight departure from a normal distribution. The critical areas in the upper tail were estimated from the table of areas of the normal distribution. It was found, for example, that in about five percent of such MDS solutions, the largest root would represent in excess of 31% of the total variance.

The already mentioned additional 15 runs were used to investigate the stability of Gamma when the number of judges changed from 20 to 80. No change was apparent. The mean Gamma was consistently in the vicinity of 25%.

An additional ten runs were run with 35 judges and only 8 stimulus objects. Changes in the number of stimulus objects influenced the distribution of Gamma considerably. This was also the case, it is recalled, with Alpha. This is to say that the size of the largest characteristic root, whether measured in absolute or relative terms, is considerably influenced by the number of stimulus objects observed.

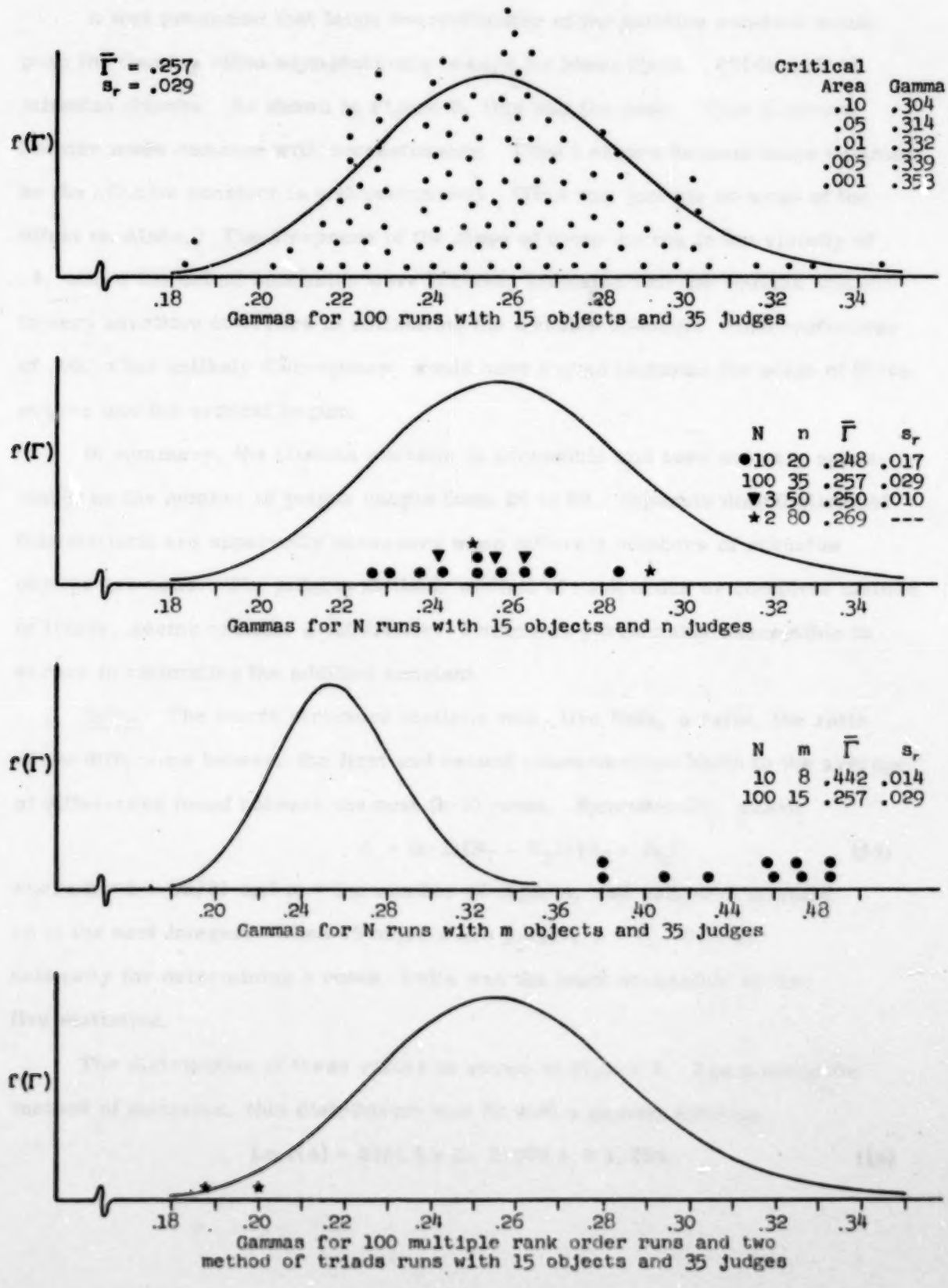

$\bar{\Gamma}$ = .257
s_r = .029
f(Γ)
Critical Area Gamma
.10 .304
.05 .314
.01 .332
.005 .339
.001 .353
.18 .20 .22 .24 .26 .28 .30 .32 .34
Gammas for 100 runs with 15 objects and 35 judges
N n $\bar{\Gamma}$ s_r
●10 20 .248 .017
100 35 .257 .029
▼3 50 .250 .010
★2 80 .269 ---
Gammas for N runs with 15 objects and n judges
N m $\bar{\Gamma}$ s_r
10 8 .442 .014
100 15 .257 .029
.20 .24 .28 .32 .36 .40 .44 .48
Gammas for N runs with m objects and 35 judges
Gammas for 100 multiple rank order runs and two method of triads runs with 15 objects and 35 judges

Fig. 7. Distribution of Gamma statistics for random input runs.
--Each of the four curves above represents the same 100 reference runs.

It was presumed that large overestimates of the additive constant would push the Gamma value asymptotically toward its lower limit, .07143 with 15 stimulus objects. As shown in Figure 8, this was the case. Type II errors become more common with overestimates. Type I errors become more common as the additive constant is underestimated. (This was just the reverse of the effect on Alpha.) The steepness of the slope of these curves in the vicinity of .5, where the actual estimates were located, indicates that the Gamma statistic is very sensitive to errors in estimating the additive constant. Underestimates of .05, a not unlikely discrepancy, would have moved Gammas for some of these curves into the critical region.

In summary, the Gamma statistic is accessible and does not vary appreciably as the number of judges ranges from 20 to 80. Separate distributions of this statistic are apparently necessary when different numbers of stimulus objects are used. The judging method, method of rank order or complete method of triads, seems to make a difference. Gamma is particularly susceptible to errors in estimating the additive constant.

Delta. The fourth structure statistic was, like Beta, a ratio, the ratio of the difference between the first and second characteristic roots to the average of differences found between the next (k-2) roots. Symbolically, where

$$\Delta = (k-2)(R_1 - R_2)/(R_2 - R_k) \qquad (14)$$

where $k = 2 + (m/3)$ and m = the number of objects, and with m/3 rounded up to the next integer. When 15 objects are judged, k = 7. With the necessity for determining k roots, Delta was the least accessible of the five statistics.

The distribution of these values is shown in Figure 9. Again using the method of moments, this distribution was fit with a gamma function.

$$\text{Ln } f(\Delta) = 2781 \text{ Ln } \Delta - 2.075\,\Delta + 1.238. \qquad (15)$$

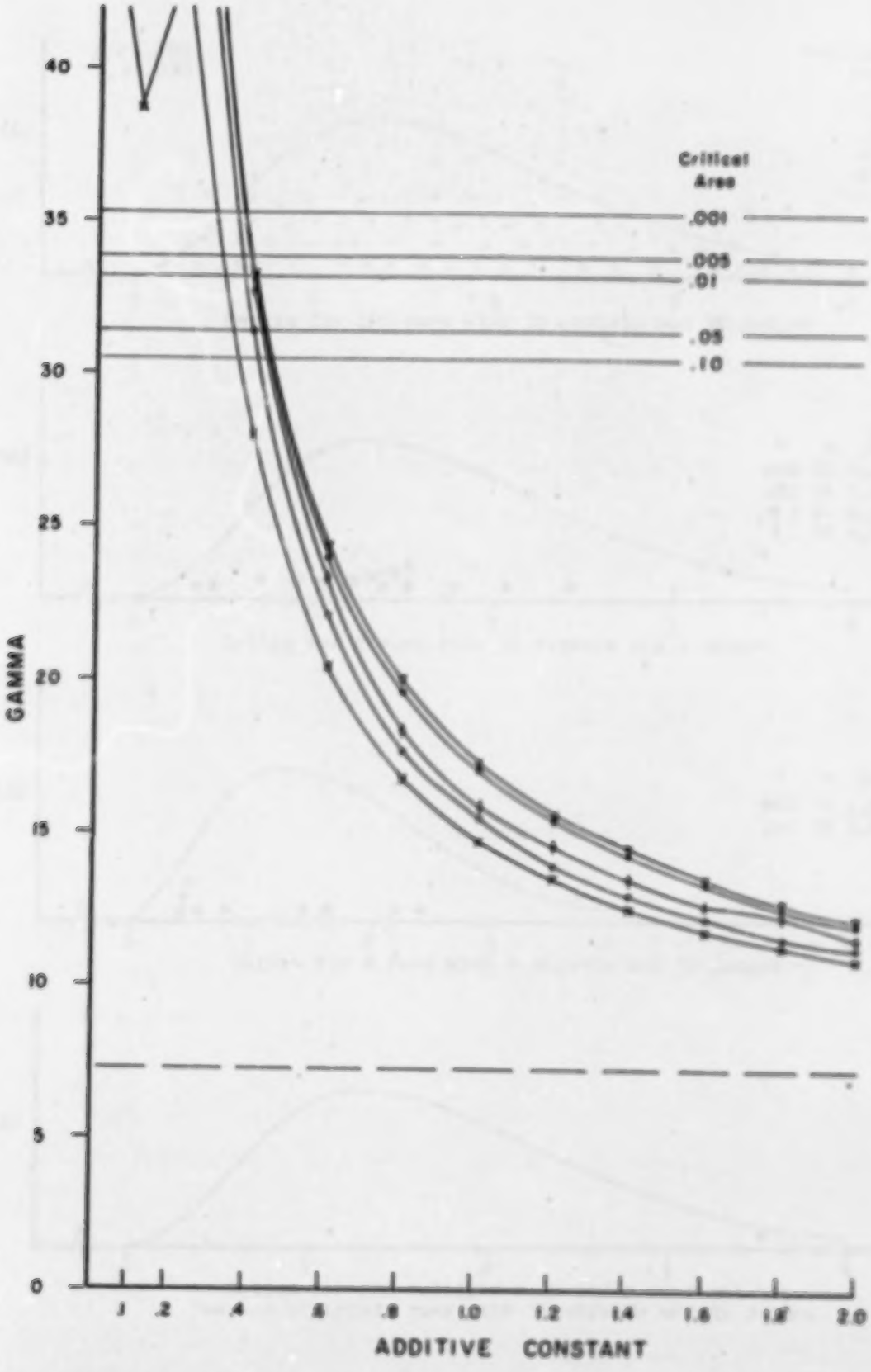


Fig. 8. Gamma as a function of the additive constant estimate for five sets of random data.

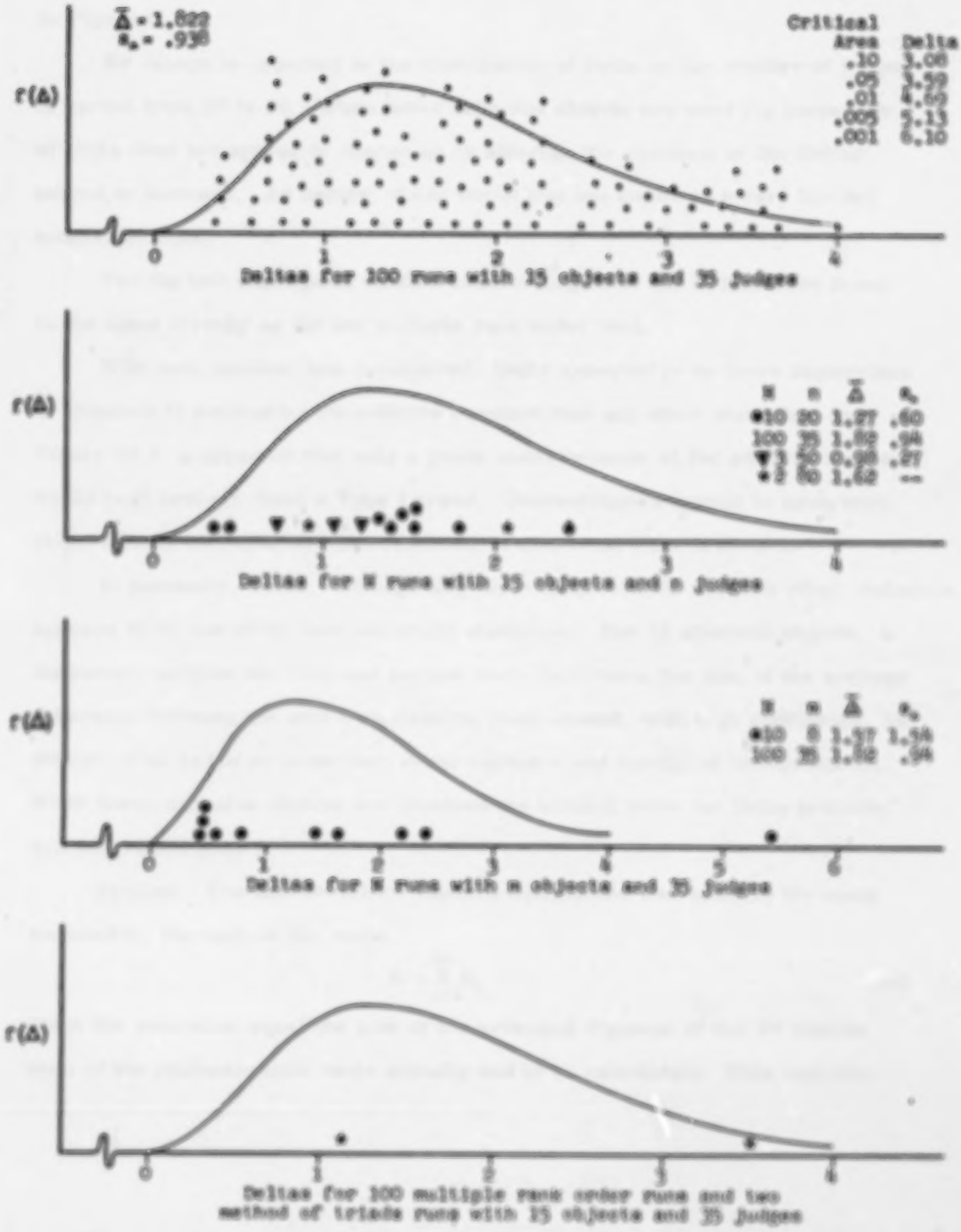


Fig. 9. Distribution of Delta statistics for random input runs.
--Each of the four curves above represents the same 100 reference runs.

From this function the critical regions were estimated and have been reported in Figure 9.

No change is apparent in the distribution of Delta as the number of judges is varied from 20 to 80. When fewer stimulus objects are used the mean size of Delta does not appear to change much although the variance of the Deltas seems to increase. As before, these variations are based on a very limited number of runs.

For the two scalings of random triadic judgments the Deltas were found in the same vicinity as for the multiple rank order data.

With only random data considered, Delta appeared to be more impervious to blunders in estimating the additive constant than any other statistic. In Figure 10 it is apparent that only a gross underestimate of the additive constant would have brought about a Type I error. Overestimates appear to have very slight effects on Delta, seldom resulting in additional Type II errors.

In summary, Delta, although slightly less accessible than the other statistics, appears to be one of the best structure statistics. For 15 stimulus objects, a difference between the first and second roots four times the size of the average difference between the next 5 successive roots should, with high confidence, be thought of as based on more than error variance and worthy of interpretation. When fewer stimulus objects are involved the critical point for Delta probably will be greater than 4.

Epsilon. The last structure statistic considered was actually the most accessible, the sum of the roots.

$$E = \sum_{j=1}^{m} \mu_j \qquad (14)$$

Since the sum must equal the sum of the principal diagonal of the B^m matrix, none of the characteristic roots actually had to be calculated. This statistic

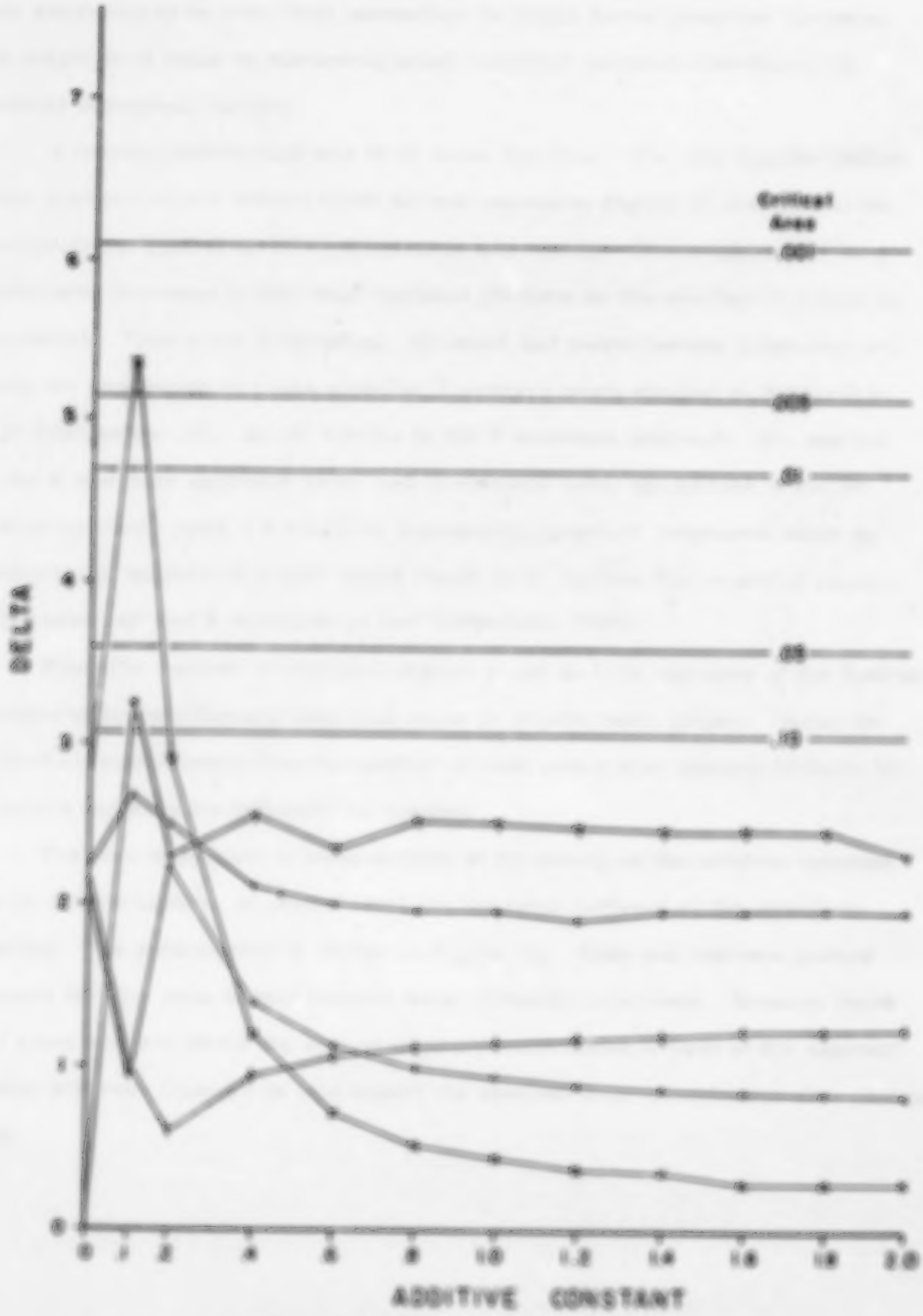


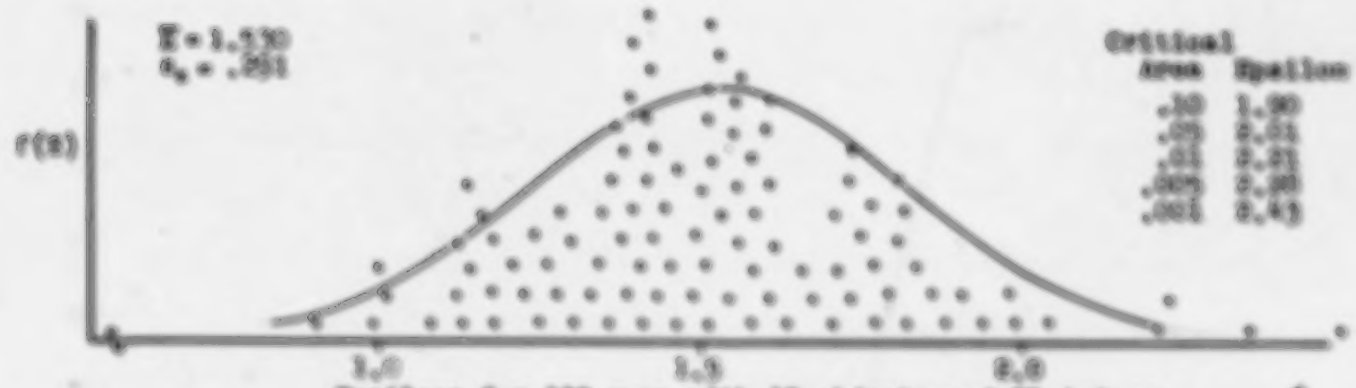
Fig. 10. Delta as a function of the additive constant for five runs with 15 objects and 35 judges.

was presumed to be relatively insensitive to single factor nonerror variance, but might be of value in discerning small nonerror variance distributed on several orthogonal factors.

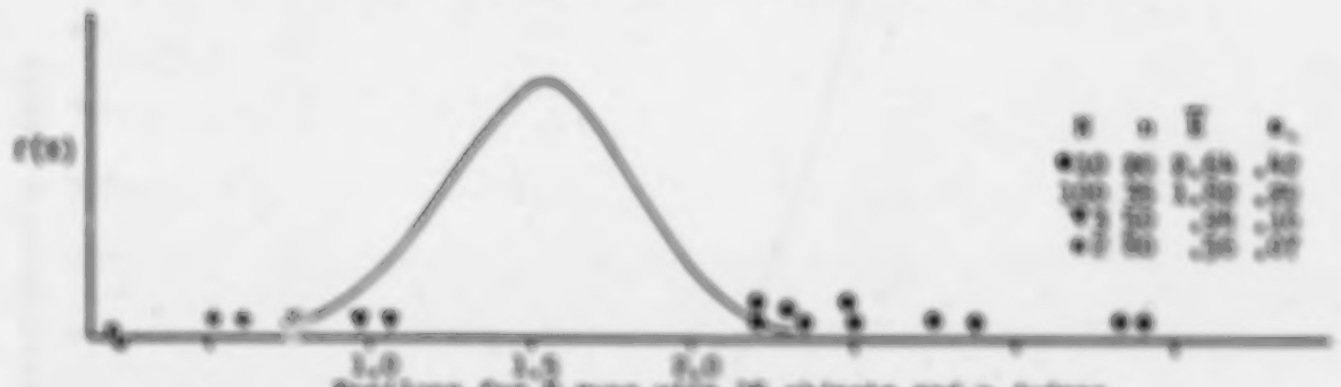
A normal distribution was fit to these Epsilons. The fact that the scales were changed for the bottom three normal curves in Figure 11 testifies to the fluctuation in Epsilon as MDS parameters are varied. There appears to be a systematic decrease in this total variance statistic as the number of judges is increased. This is not surprising. As more and more random judgments are made the proportion of times stimulus F appears more similar to H than G is to H approaches .50. As all entries in the P matrices approach .50, entries in the R matrices approach zero, and in ultimate turn, all entries in the R^* matrix approach zero. It would be expected that random judgments made by a very large number of judges would result in an Epsilon that would be essentially zero. (P and R matrices as per Torgerson, 1958).

When the number of stimulus objects is set at 9 the variance of the matrix appears to be considerably less than when 13 objects were judged. Using the method of triads rather than the method of rank order also appears (Figure 11) to have a constrictive influence on Epsilon.

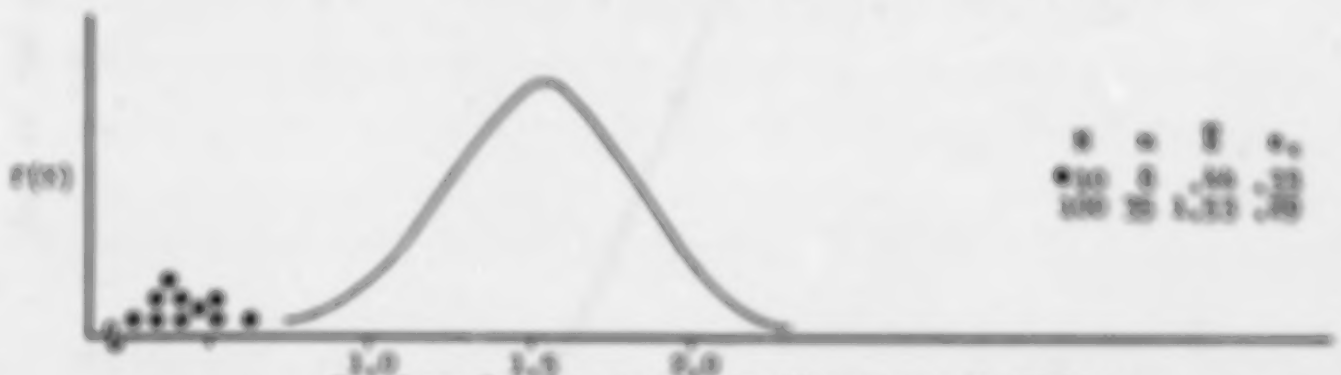
The size of Epsilon is considerably at the mercy of the additive constant. Large overestimates, of course, will set the total variance of the structure soaring. The relationship is shown in Figure 12. Only one run was plotted because the five runs in this respect were virtually coincident. Because there was some concern about the size of negative roots found in each of the approximately 200 runs covered by this report the absolute sum of roots was also plotted here.



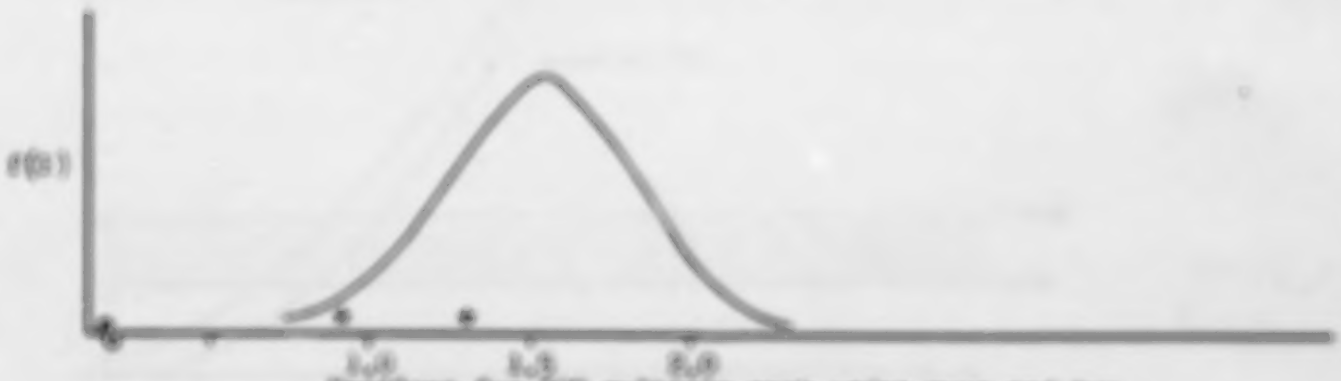
Epsilons for 100 runs with 15 objects and 25 judges



Epsilons for 9 runs with 15 objects and 9 judges



Epsilons for 9 runs with 9 objects and 25 judges



Epsilons for 100 multiple rank order runs and two samples of single runs with 15 objects and 25 judges

Fig. 13. Distribution of Epsilon statistics for random input runs. Each of the four curves above approximate the same 100 reference runs.

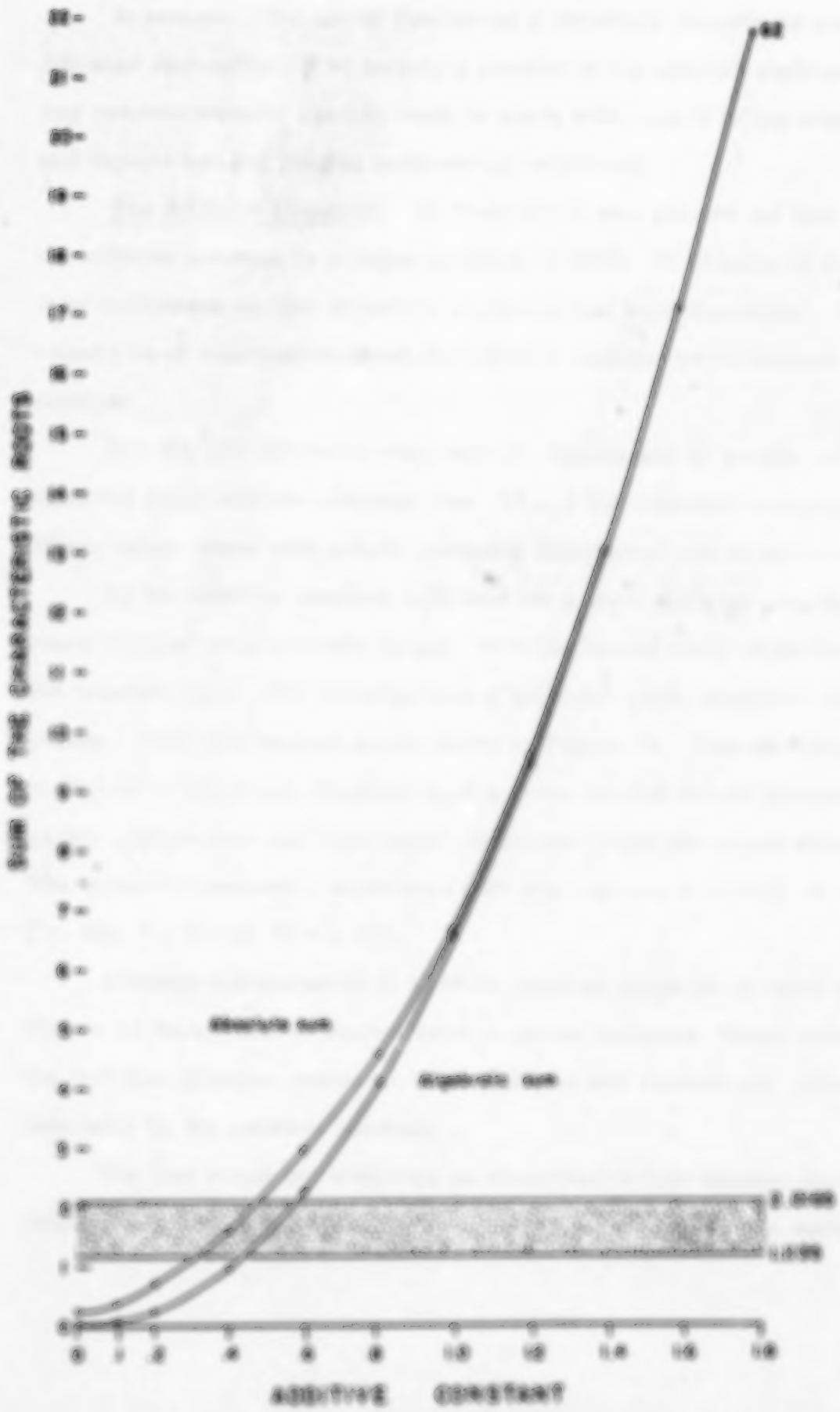


Fig. 12. The algebraic sum (Epsilon) and the absolute sum of ranks as a function of different values of the additive constant.

In summary the use of Epsilon as a structure statistic is quite limited. Although accessible, it is largely a product of the additive constant estimate. Any consideration of Epsilon must be made with regard to the numbers of judges and objects and the judging methodology employed.

The Additive Constant. In Chapter I it was pointed out that estimation of the additive constant is a major problem in MDS. In Chapter II the influence of poor estimation on five structure statistics has been discussed. Several additional bits of information about the additive constant were deemed worthy of mention.

For the 100 reference runs with 15 objects and 35 judges, and random input the mean additive constant was .50 and the standard deviation was .04. These values were very nearly normally distributed and ranged from .39 to .64.

As the additive constant estimate for a given solution was increased, the ranks became progressively larger, with the larger ranks expanding faster than the smaller ones. The relative size of adjacent ranks remained remarkably steady. This information is portrayed in Figure 13. The solutions represented in Figure 13 are those obtained from a given random error structure (Set D10) as the additive constant was varied above and below the actual estimate, .4000. The structure statistics associated with this run are A = .413, B = 1.332, Γ = .341, A = 2.948, E = 1.472.

Although the variation in additive constant estimate is much greater in Figure 13 than would be encountered in actual analyses, these curves reflect the fact that absolute statistics such as Alpha and Epsilon are influenced considerably by the additive constant.

The five structure statistics as described in this chapter were examined only in terms of random data input. Further examination was made with

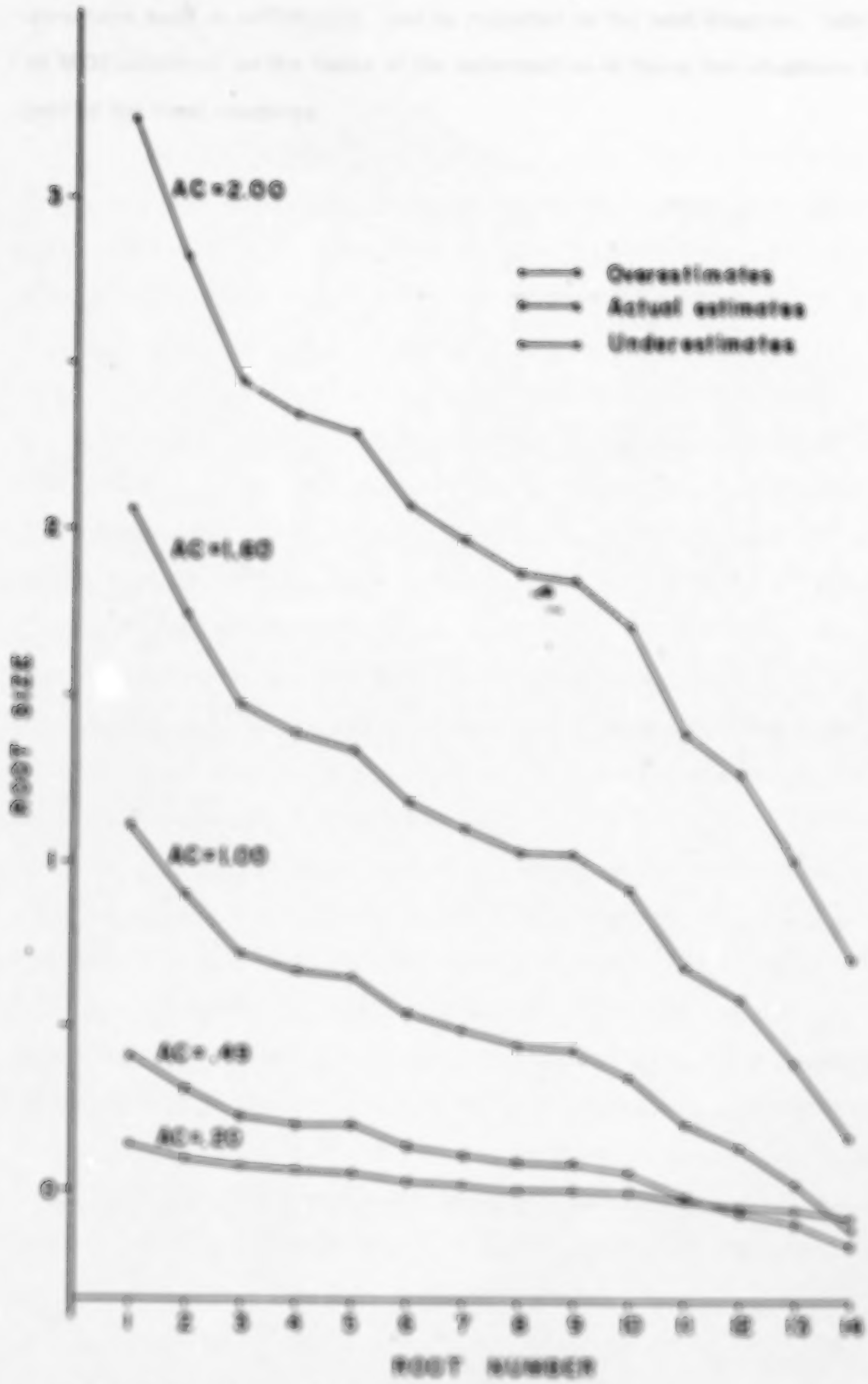


Fig. 13. Size of characteristic roots with various estimates of the additive constant for one set of ESS data.

structure built in artificially, and is reported in the next chapter. Interpretation of MDS solutions on the basis of the information of these two chapters is the subject of the final chapters.

CHAPTER III

ARTIFICIAL STRUCTURE

The next step in the evaluation of the structure statistics which have been defined involved artificially built-in structure, nonerror structure that was imposed on the random structure. It was believed that the five statistics would differ with regard to their sensitivity to small nonerror variance. Therefore, nonerror variance was added to the structure until the structure statistics were inflated well beyond the random input values observed in the preceding chapters.

Artificial structure was obtained by adding, for each stimulus object at the time of each assessment, a designated "true" scale value to a randomly drawn "perception error." Let us assume that objects A, B, and C have true scale values of 20, 30, and 50. Three perception errors are drawn, in order, 4, -12, and -6. The perceived scale values for A, B, and C would be the algebraic sums: 24, 18, and 44. The assessment question, "Is A or is B more like C?" is answered, "A." As perception errors become small compared to differences in "true" scale values the nonerror variance becomes more prominent and artificial structure has been built in.

With no artificial dimensionality all objects are assigned the same true scale value. The perception error is selected randomly from a specified distribution of errors. This distribution of perception errors is the basis for Thurstone's discriminal dispersion. In this study the distribution, as usual, was considered to be normal. If these perception errors are given normal deviate values, the mean perception error is zero and the standard deviation unity. We will use units of this same scale to specify the differences between "true" scale values of the stimulus objects. That is to say, the differences between "true" scale values will be expressed in units defined by the discriminal dispersion.

When one dimension was desired, the "true" scale values assigned to the objects were the same values for all judges. Judgments differed from judge to judge, of course, due to the random perception error included. When two or more orthogonal, artificial dimensions were desired the judges were divided into two or more groups. Within a group the same "true" scale values were assigned to the objects for all judges, but the scale value assignment for any two groups was uncorrelated.

Scale Values. Two different unidimensional patterns of "true" scale values were utilized, one with two clusters of coincident objects with a given distance between clusters and one with objects scattered at equal intervals along the continuum. These will be referred to as the cluster pattern and the scatter pattern of stimulus objects. A common distance scale was obtained by converting cluster distances and scatter intervals into average interpoint distances (AIPD). For the cluster method,

$$AIPD = \frac{2m_1m_2d}{m(m-1)} \tag{27}$$

and for the scatter method,

$$AIPD = \frac{(m^3 - m)\,d}{6} \tag{28}$$

where m is the number of objects

m_1 is the number of objects in cluster 1

m_2 is the number of objects in cluster 2

and d is the standard deviation distance between clusters or interval between adjacent objects, as appropriate.

It was believed that for a given AIPD any observed arrangement of stimulus objects yields results somewhere between the limits established by these two patterns.

With 15 objects, as most commonly used in this study, if the clusters are one standard deviation distance apart, the AIPD will be about .5 sigma; and if the objects are scattered at intervals of .10 the AIPD is also about .5 sigma. The relationship, of course, is linear.

The Sort. Two sequences of random normal deviates were used extensively. These sequences are designated here as Sort A and Sort H. The error variance or error structure for all runs with a given sort is constant. A different sort means a different error structure.

The Structure Statistics and AIPD. The effect of greater and greater unidimensional nonerror variance on the structure statistics is shown in Figures 14-18. For each point plotted a solution involving 15 objects and 35 judges was obtained. Each curve is based upon a certain statistic, a certain sort, and a certain pattern of stimulus objects. Level of significance lines, as determined in the previous chapter, are shown.

In Figures 14-18 it is obvious that after the nonerror variance becomes as prominent as the error variance (after the AIPD reaches a certain value) additional increase in AIPD results in marked increase in the structure statistics. With still greater increase in AIPD the functions do level off, however. The distances between objects becomes so great that the perception errors are of no consequence in the judging. The structure statistics reach a limit because they are a function of the certainty of judgment rather than a function of the actual differences among objects.

Comparison of Statistics as to Sensitivity. To simplify the comparison of the curves for the different statistics as they cross the level of significance lines, Figure 19 is presented. The four curves of each of the preceding five figures have been combined into a single curve segment. Considerable rounding and interpolation was involved in preparation of this figure. To the extent that

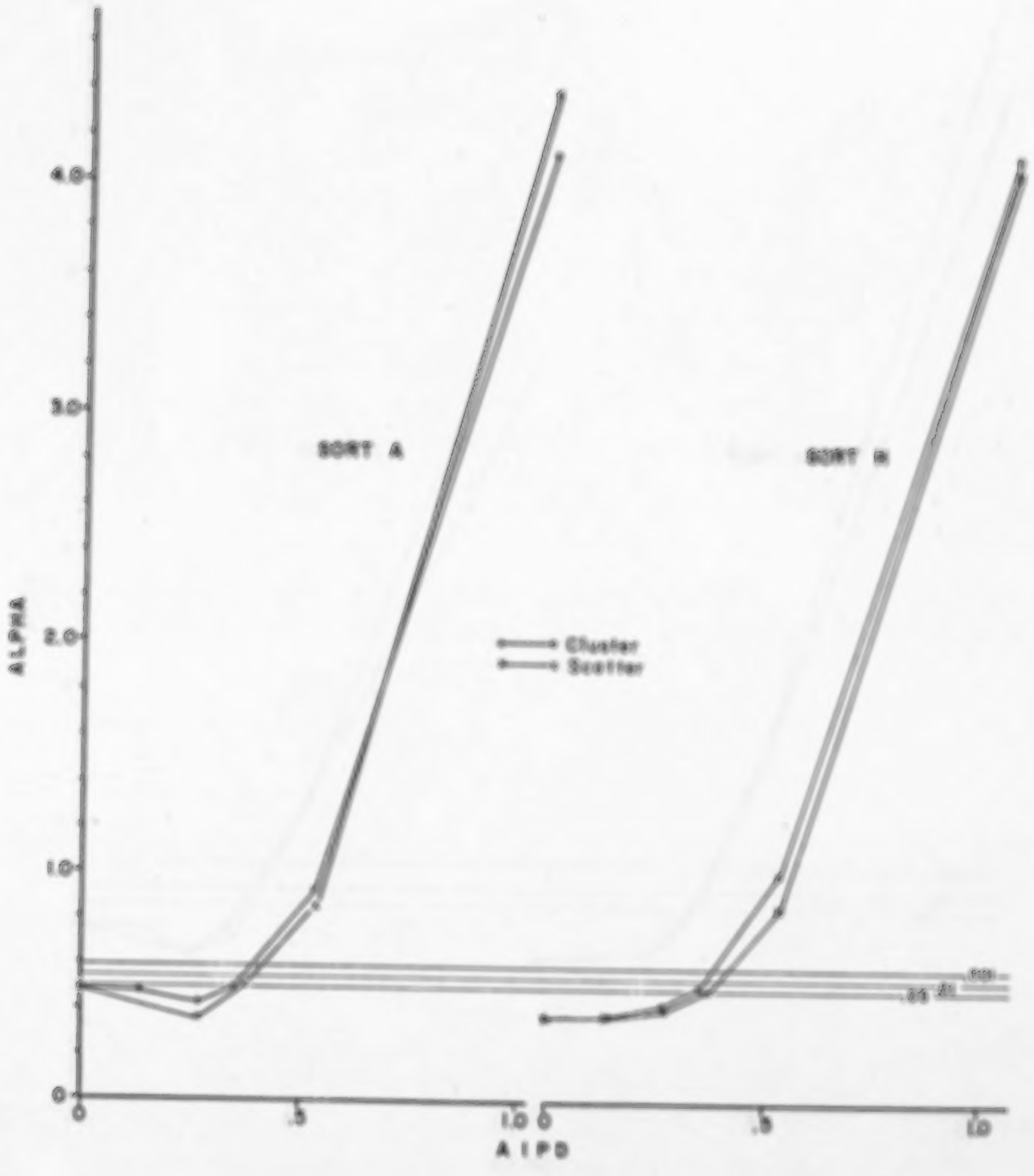


Fig. 14. Alpha as a function of average interpoint distance with one nonerror dimension.

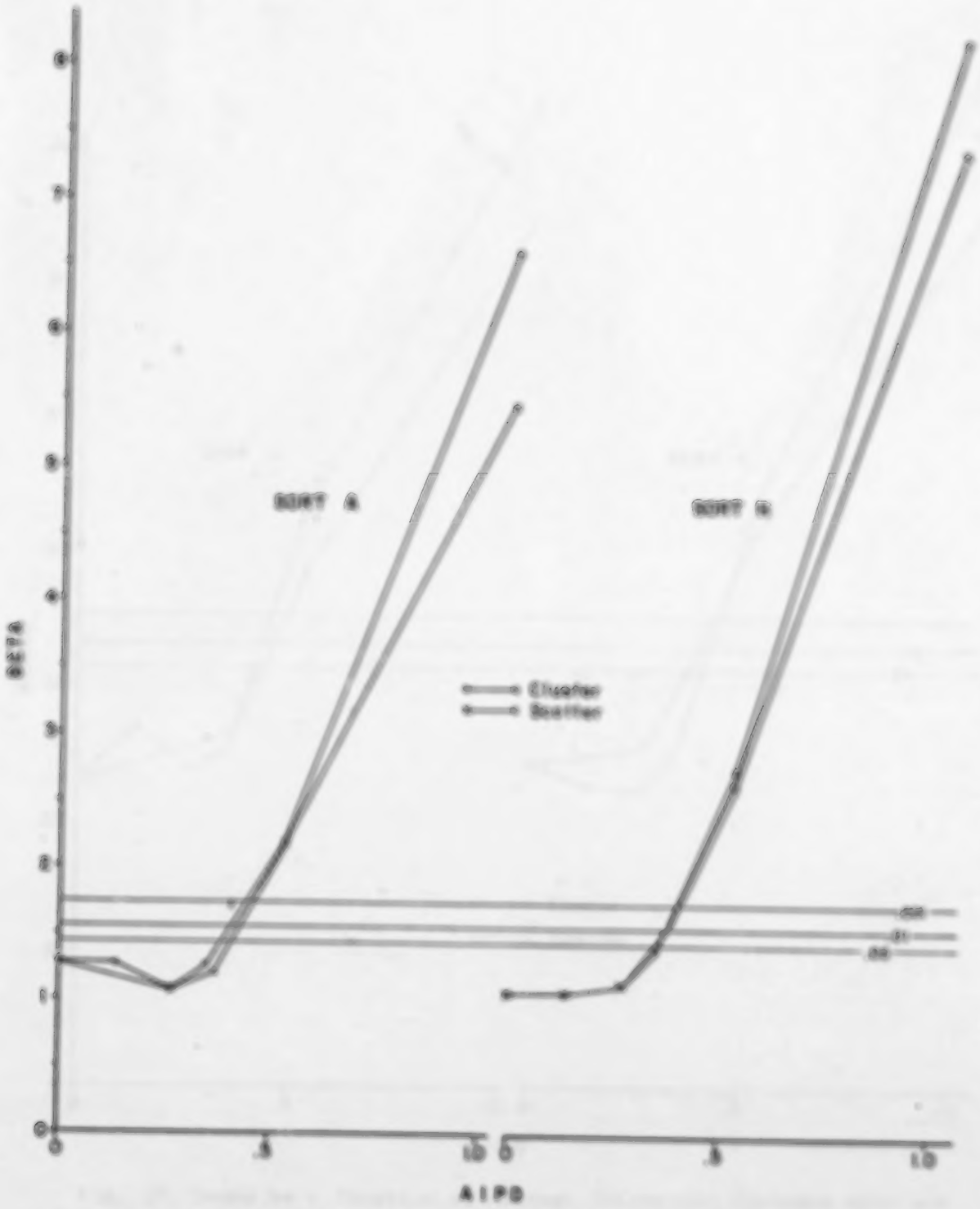


Fig. 15. Beta as a function of average interpoint distance with one nonerror dimension.

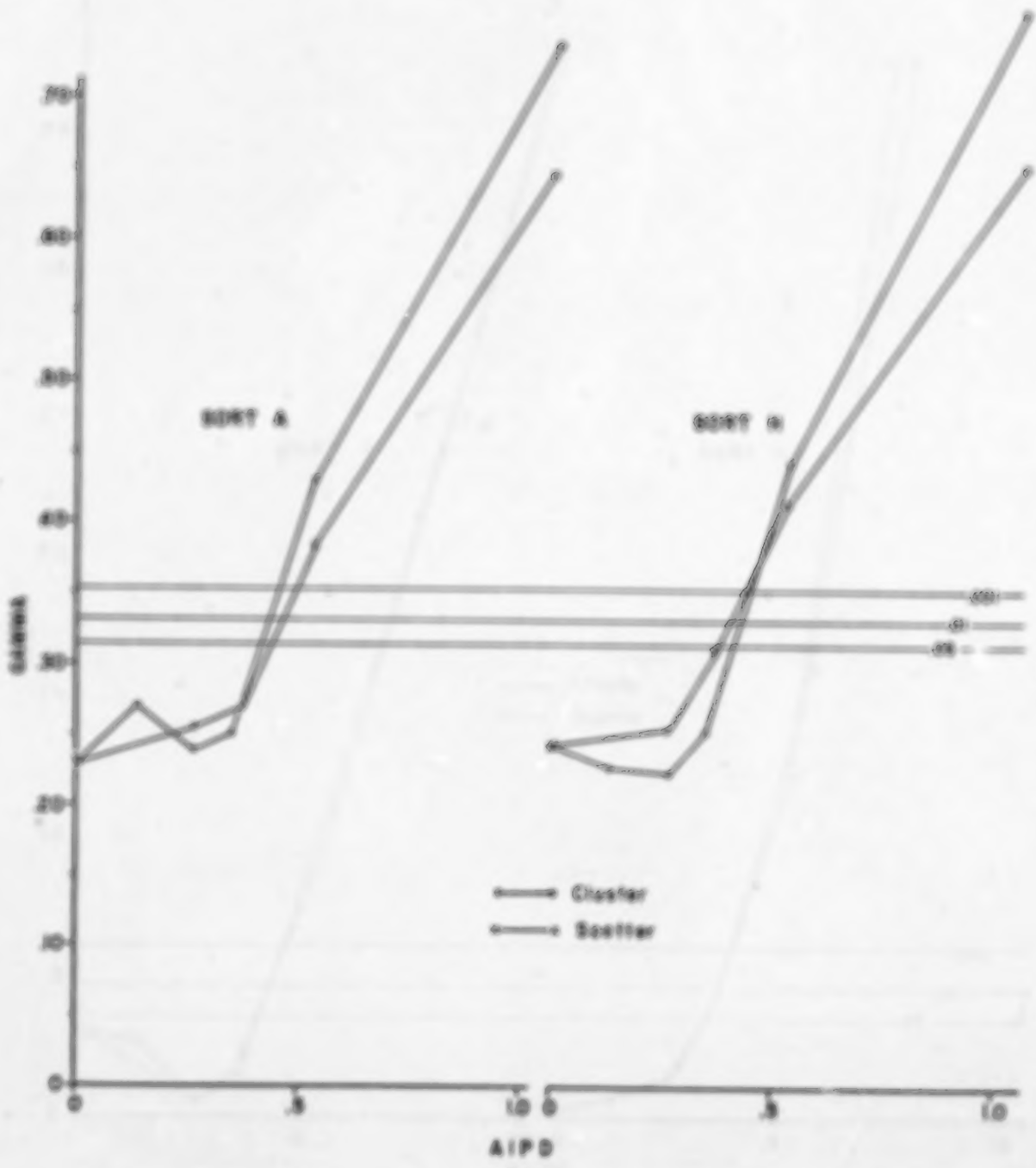


Fig. 16. Gamma as a function of average interpoint distance with one noncrror dimension.

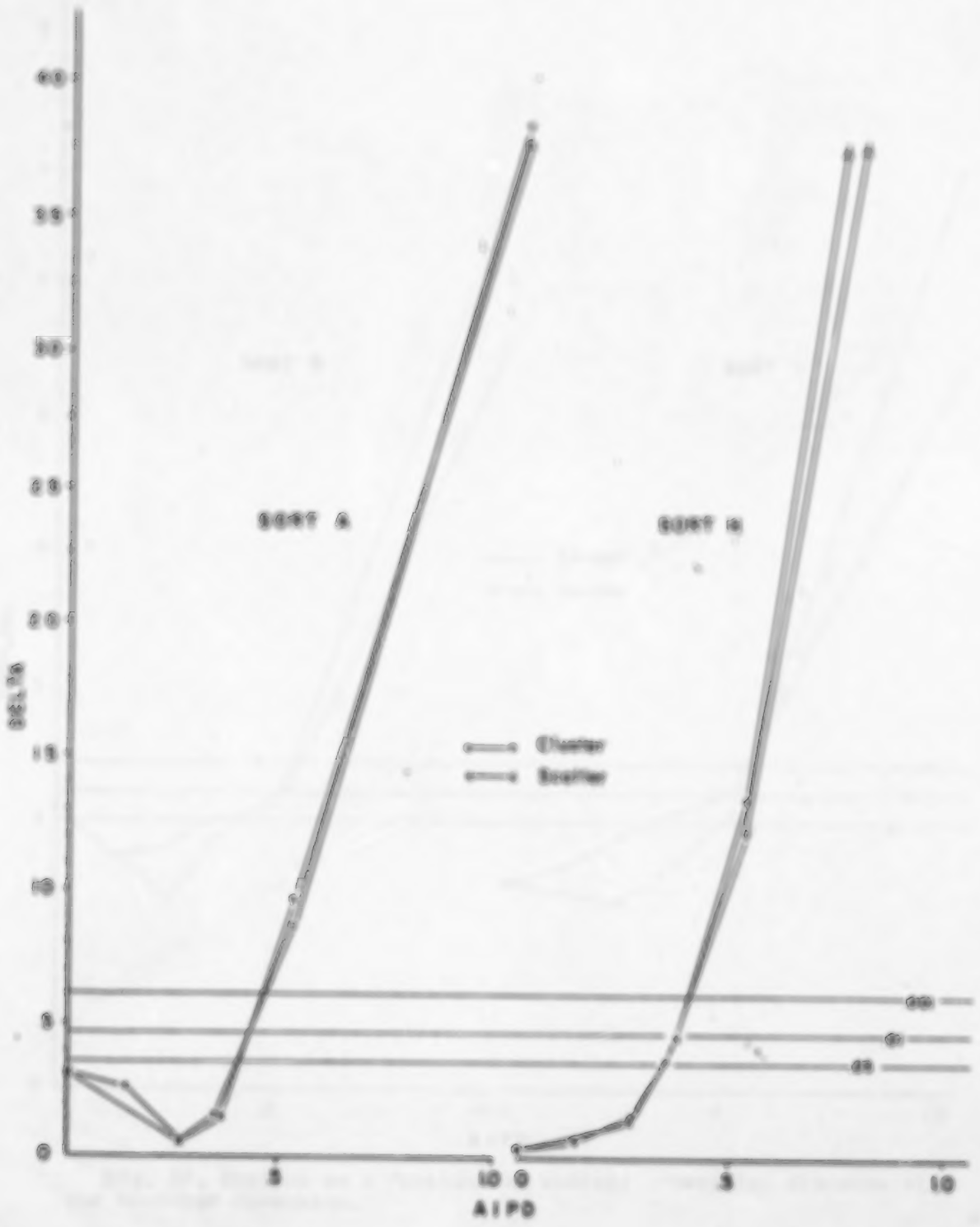


Fig. 17. Delta as a function of average interpoint distance with one nonerror dimension.

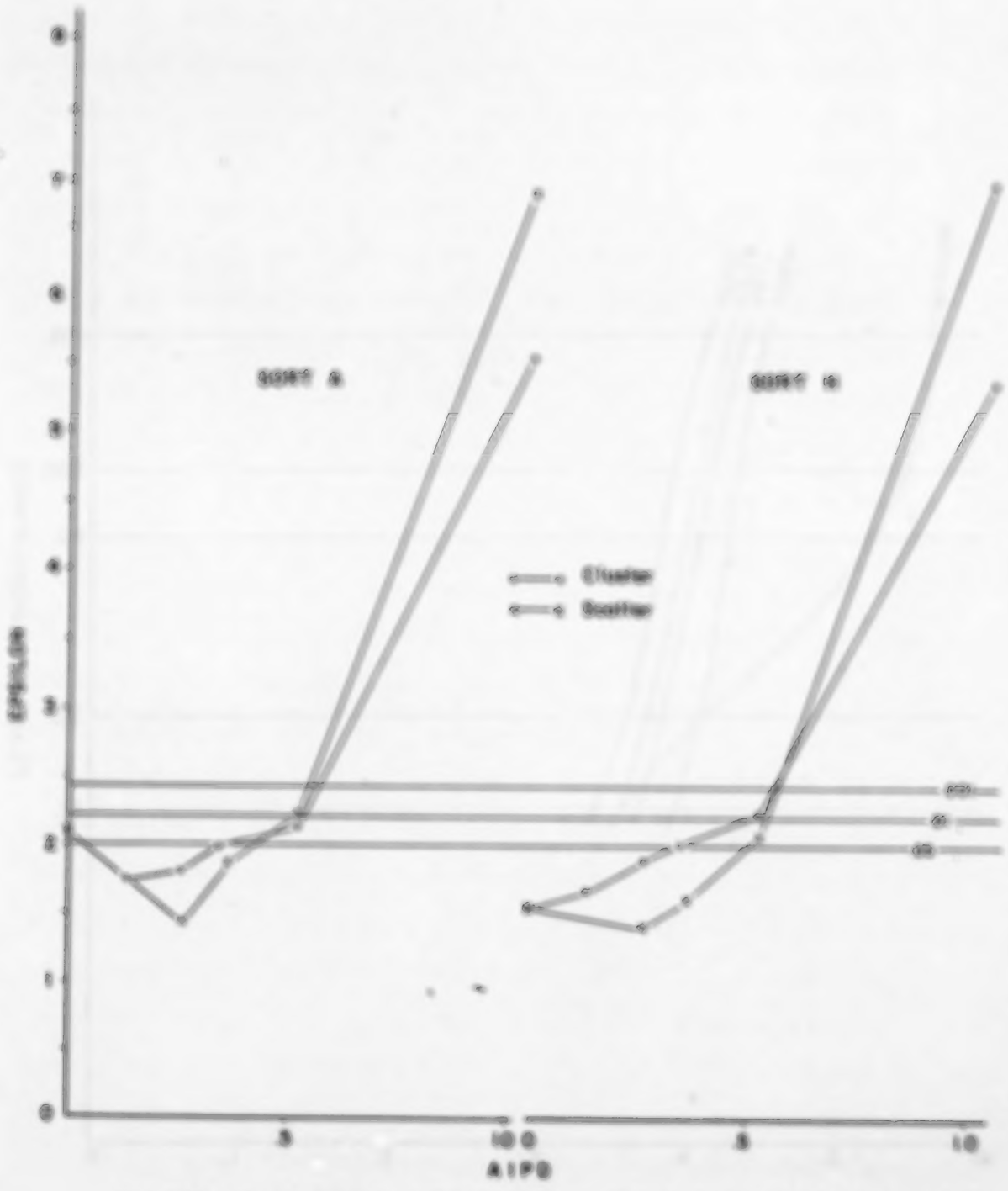


Fig. 10. Epsilon as a function of average interpoint distance with one noncrror dimension.

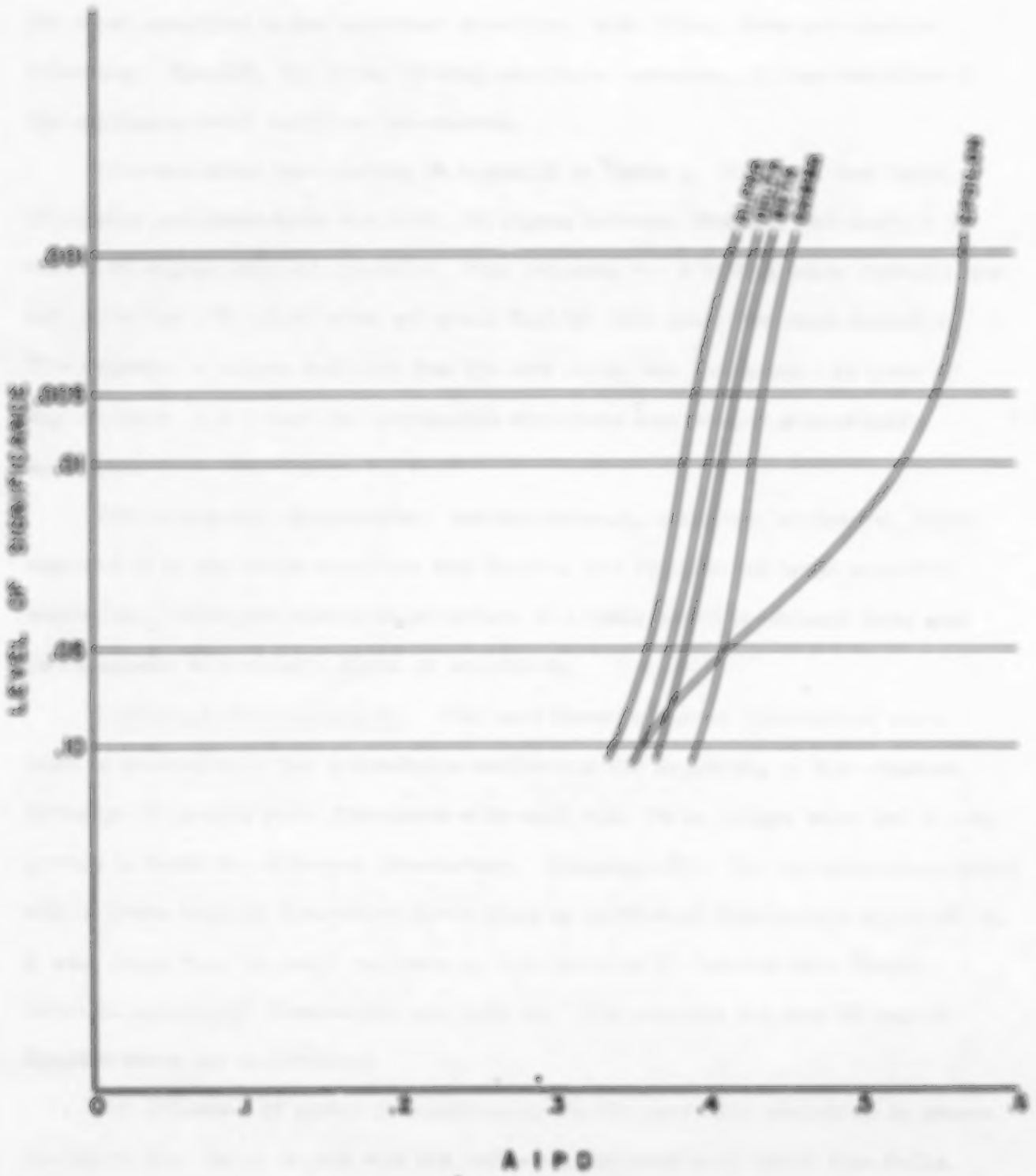


Fig. 19. A graphic comparison of five structure statistics as one dimensional structure is built up.

comparisons is justified, the three statistics based primarily on the size of the largest root and Delta behave in a very similar fashion. Alpha appears to be the most sensitive to the nonzero structure, with Delta, Beta and Gamma following. Epsilon, the index of total structure variance, is less sensitive to the unidimensional variance increments.

Corroborative information is revealed in Table 5. To make that table, 50 cluster solutions were run with α.05 sigma between clusters and another 50 with 1.00 sigma between clusters. The columns for α in that table indicate how far above the α.05 cutoff point we would find the new scree for each statistic. The negative α values indicate that the new scree was below the α.05 level of significance, i.e., that the nonspatial structure was not yet statistically apparent.

With minimally discernible, unidimensional, nonzero structure, Alpha appears to be the most sensitive and Gamma and Epsilon the least sensitive statistics. When the nonzero structure is a little more prominent Beta and Delta appear to overtake Alpha in sensitivity.

Additional Dimensionality. Two and three nonzero dimensions were built in according to the procedures outlined at the beginning of this chapter. Although 50 judges were simulated with each run, three judges were put in subgroups to build the different dimensions. Consequently, the variance associated with a given built-in dimension diminishes as additional dimensions are built in. It was found that the total variance as represented by Epsilon also diminishes as additional dimensions are built in. The reasons for this change in Epsilon were not understood.

The influence of added dimensionality on the structure statistics is shown in Figure 20. Beta, which was not included, behaved very much like Delta;

TABLE 1

The Sensitivity of Five Statistics as Revealed by the Distance Each is Moved into the Critical Region by Building-in Unidimensional Nonerror Structure

	With only error structure ($N = 100$ runs)			With artificial structure added $\Delta STD = .05$ ($N = 10$ runs)				With artificial structure added $\Delta STD = .10$ ($N = 10$ runs)			
	$\overline{X}_0$	s_0	$\overline{X}_{.05}$	$\overline{X}_1$	s_1	s_Δ	w_1	$\overline{X}_0$	s_0	s_Δ	w_Δ
Alpha	.007	.003	.008	.007	.001	.006	.02	.070	.009	.009	7.70
Beta	1.005	.007	1.021	1.000	.070	.015	.3	5.050	.000	1.090	10.10
Gamma	.007	.010	.016	.000	.002	-.010	-.00	.000	.007	.010	0.00
Delta	1.01	.000	2.10	2.70	1.00	.0	.01	10.01	0.00	0.10	0.00
Epsilon	1.000	.005	2.000	1.700	.070	-.010	-.70	0.010	.010	.000	.70

$s_\Delta = \overline{X}_1 - \overline{X}_{.05}$

$w_\Delta = (s_\Delta)/(s_0)$ = number of standard deviation distances above the .05 level to the statistics new mean

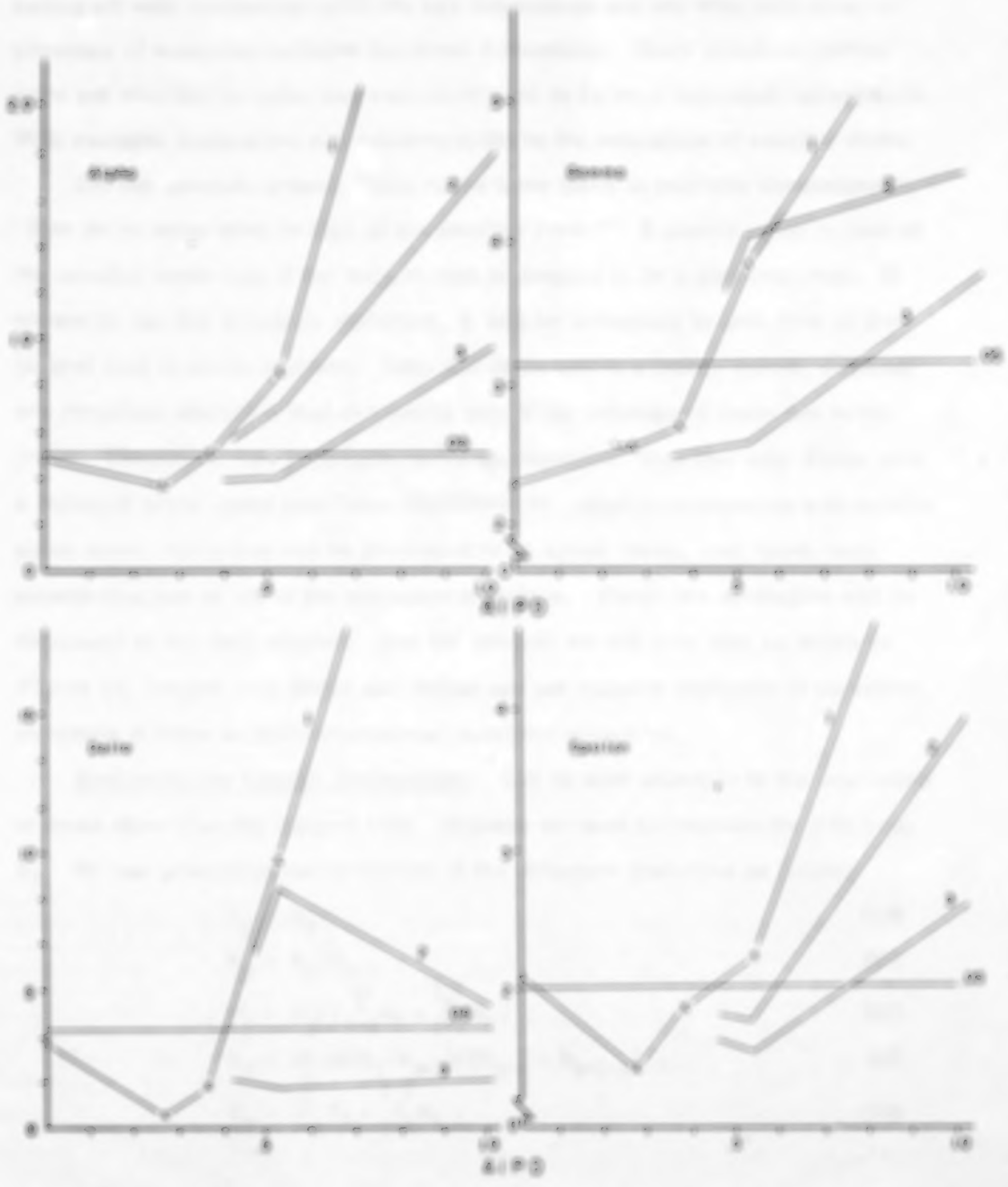


Fig. 30. Structure statistics for the largest root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in.

falling off with increasing $\Delta^2(\theta)$ for two dimensions and not even indicating the presence of nonzero optimum for three dimensions. More complete curves were not obtained because this was not thought to be very important information. With multiple dimensions our concern shifts to the evaluation of smaller roots.

But the question arises, "How do we know there is multiple dimensionality?" "How do we know when to look at the smaller roots?" It makes sense to look at the smaller roots only if the largest root is deemed to be a nonzero root. If we are to use the structure statistics, it will be necessary to look first at the largest root in every problem. Bois and Dalla and to a lesser extent, Gleason, are structure statistics that are useful only if the subsequent roots are zero roots. Therefore, two strategies seem appropriate: consider only flights until a string of zero roots have been identified or start consideration with middle sized roots, roots that can be presumed to be zero roots, and work back, considering any or all of the structure statistics. These two strategies will be discussed in the final chapter. For the present we will note that as shown in Figure 10, largest root Bois and Dalla are not reliable indicants of nonzero structure if there is multidimensional nonzero structure.

Evaluating the Lesser Dimensions. Let us shift attention to the evaluation of roots other than the largest root. Suppose we want to evaluate the j-th root, θ_j. We can generalize the definition of the structure statistics as follows

$$A_j = \theta_j \quad (i)$$

$$B_j = \theta_j/\theta_{j+1} \quad (ii)$$

$$C_j = \theta_j/\left(\sum_{i=j}^{m}\theta_i - \sum_{i=j}^{j-1}\theta_i\right) \quad (iii)$$

$$D_j = \left(\frac{m-j}{m}\right)(\theta_j - \theta_{j+1})/(\theta_{j+1} - \theta_{j+2}) \quad (iv)$$

$$E_j = \sum_{i=j}^{m}\theta_i - \sum_{i=j}^{j-1}\theta_i \quad (v)$$

all of which is merely a matter of removing from consideration those nodes larger than R_j and proceeding as before.

Whether or not the distributions considered in the previous chapter are of any value in evaluating R_j will be discussed in the next chapter.

With known dimensionality, when the smallest nonzero root (root 1 with one dimension, root 2 with two dimensions, etc.) is examined, we make an expected finding: with added dimensionality it takes more nonzero variance (RMS) to move the statistic into the critical region. Of the first four statistics, Gamma appears to be the most sensitive (Figure 5). There is hardly enough variance associated with the third root with 10 objects, 10 judges and 2 artificial dimensions to assure its nonzero status when RMS is less than unity unless Gamma is used. Although Kinds, Beta, Gamma and Delta were about equally sensitive to 1 dimension structure, only Gamma seemed to maintain that sensitivity to multidimensional structure.

A question that now comes up is, "Do the error roots remain at a constant level while artificial structure is built up on the nonzero roots?"

Increased Error Variance. Given a theoretical, when multidimensional, nonzero structure is combined with error structure, that one dimension is unlikely to coincide with any of the orthogonal dimensions. If the nonzero variance is small, it should move on into split up and attach itself to several of the error dimensions. If the nonzero variance is large, it should do much to determine the location of the first root. In this case the error roots would be rearranged somewhat but would not increase in size.

Unfortunately, that is not what appears to happen in the Tungsram RMS procedure. As multidimensional nonzero structure becomes a primary determinant of the judgments, the error roots as well as the nonzero root increase

2 OF 2
ED
003060

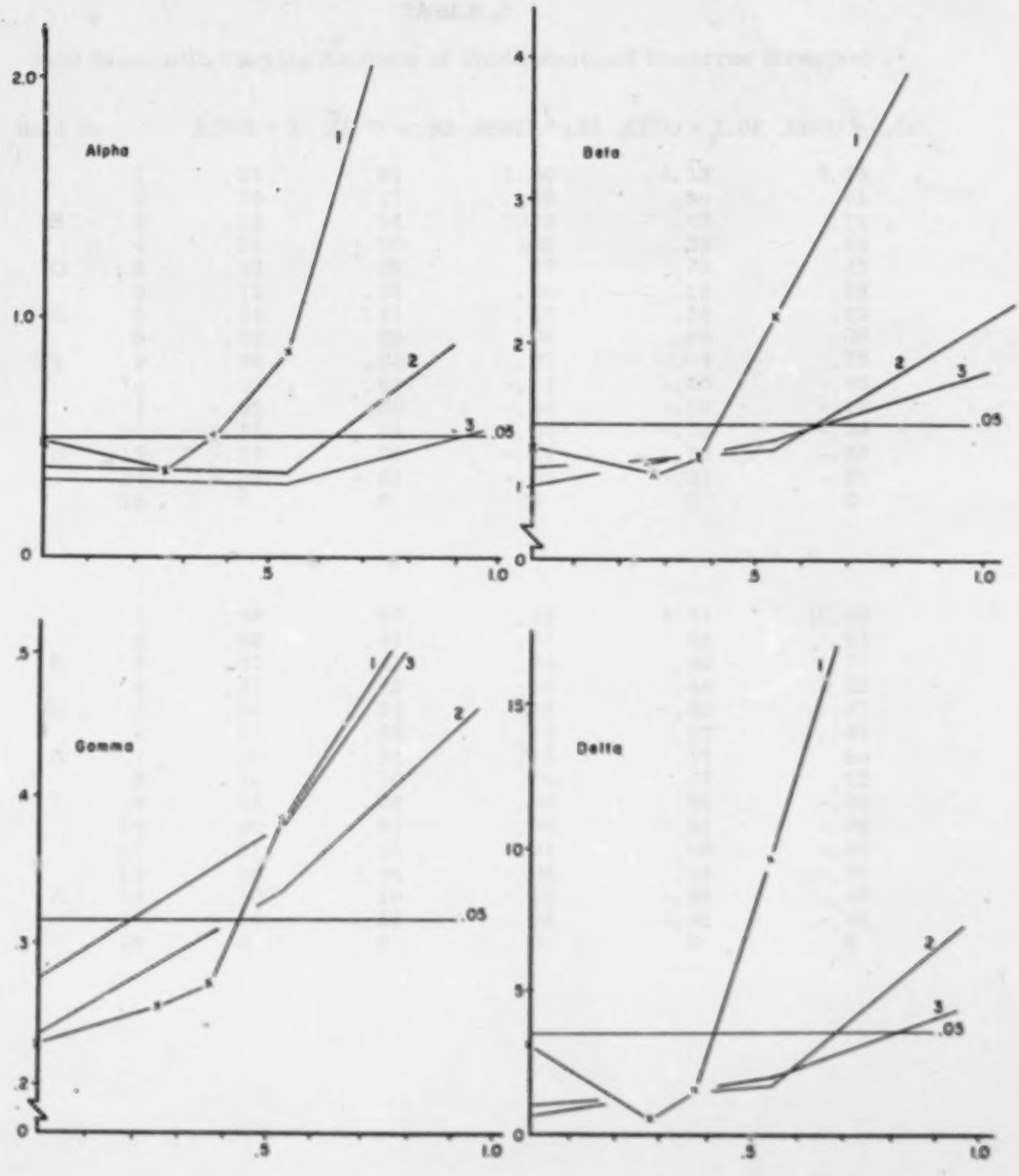


Fig. 21. Structure statistics for the smallest nonerror root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in.

TABLE 2

Root Sizes with Varying Amounts of Unidimensional Nonerror Structure

Root No.		AIPD = 0	AIPD = .35	AIPD = .53	AIPD = 1.06	AIPD = 1.60
	1	.37	.51	1.00	4.13	9.95
	2	.35	.37	.36	.50	.81
S	3	.25	.34	.35	.43	.71
	4	.24	.30	.28	.39	.44
O	5	.19	.25	.27	.29	.43
	6	.19	.23	.20	.15	.33
R	7	.14	.17	.10	.14	.23
	8	.07	.08	.08	.06	.06
T	9	.04	.06	.07	.04	.03
	10	.00	.01	-.01	-.00	-.07
	11	-.03	.00	-.04	-.08	-.10
	12	-.03	-.04	-.05	-.14	-.23
H	13	-.12	-.08	-.14	-.24	-.38
	14	-.17	-.21	-.23	-.32	-.55
	15	0	0	0	0	0
	1	.48	.50	.92	4.11	10.58
	2	.38	.41	.41	.62	.81
S	3	.33	.37	.34	.46	.71
	4	.31	.25	.26	.40	.52
O	5	.26	.24	.22	.32	.47
	6	.21	.20	.19	.27	.37
R	7	.20	.15	.12	.17	.25
	8	.12	.10	.06	.07	.12
T	9	.08	.08	.00	.04	-.00
	10	.00	.01	-.00	-.01	-.05
	11	-.01	-.03	-.01	-.06	-.10
	12	-.06	-.05	-.06	-.15	-.25
A	13	-.10	-.10	-.12	-.28	-.43
	14	-.15	-.18	-.22	-.41	-.48
	15	0	0	0	0	0

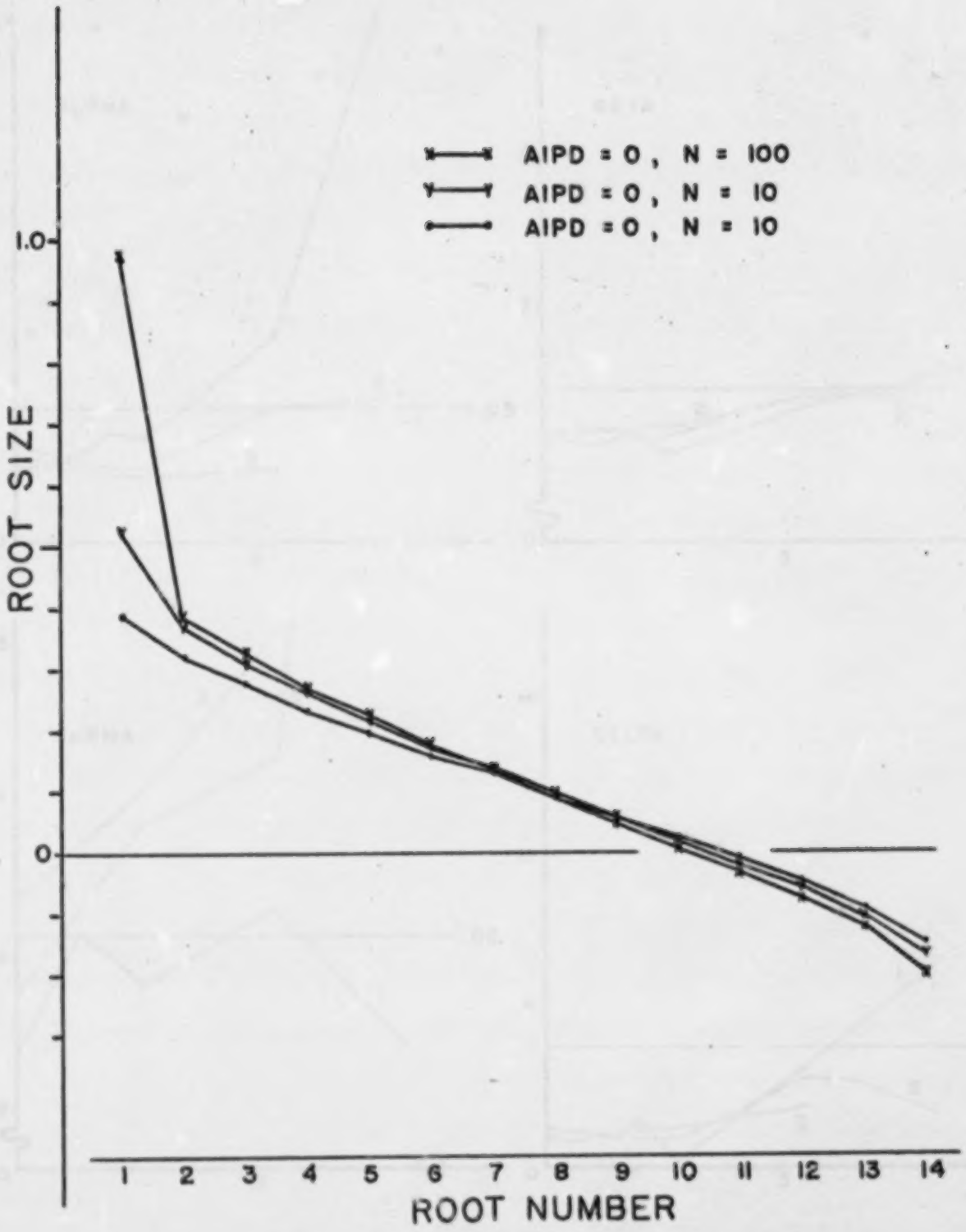


Fig. 22. Mean root sizes for 15x25 MDS runs with different amounts of nonerror structure built in.

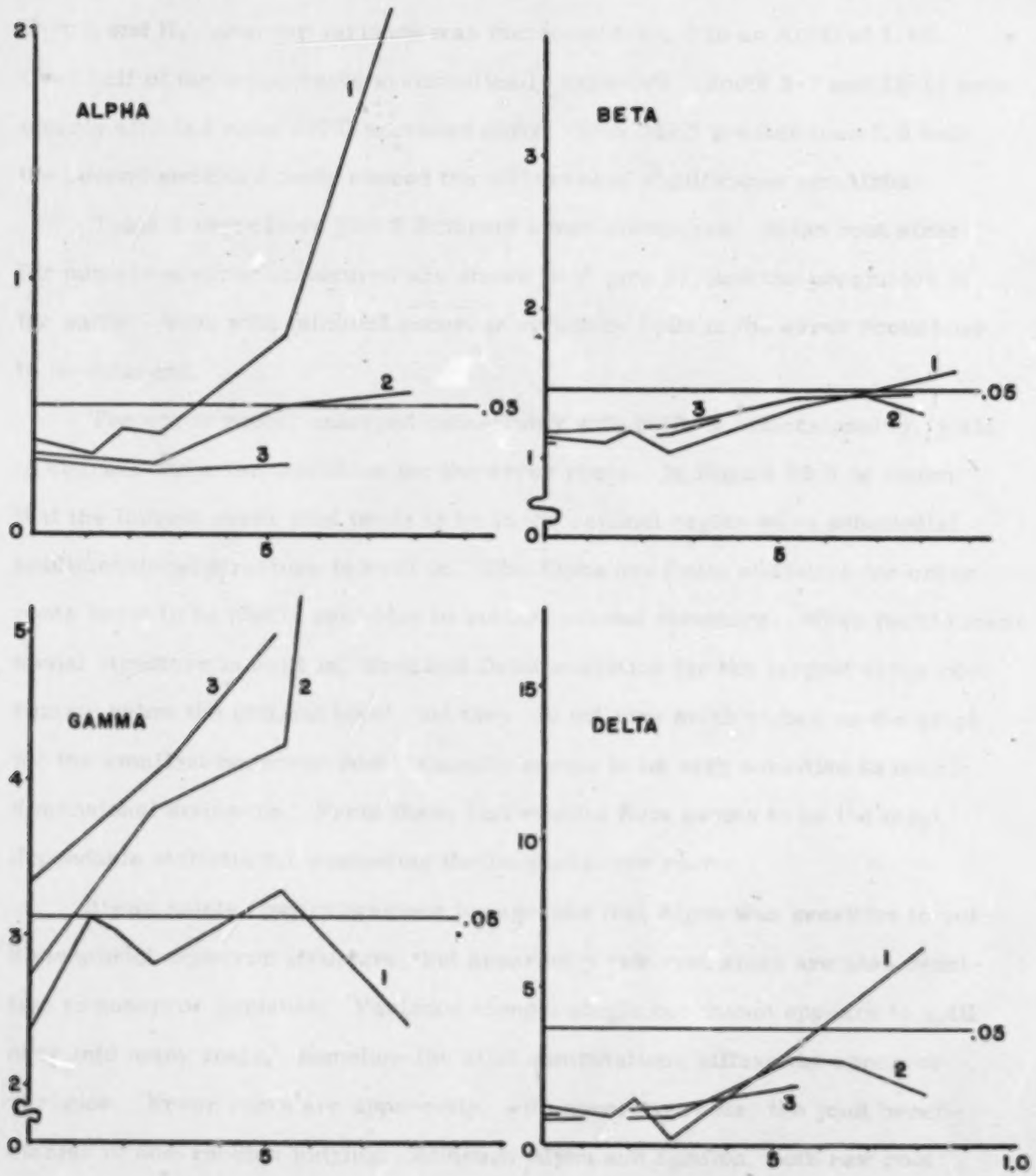


Fig. 23. Structure statistics for the largest error root as a function of average interpoint distance when 1, 2, and 3 dimensions were built in.

substantially. This can be observed in Table 2 where for two error structures, SortsA and H, nonerror variance was increased from 0 to an AIPD of 1.60. Over half of the error roots systematically expanded. Roots 2-7 and 12-14 were clearly affected when AIPD exceeded unity. With AIPD greater than 1.5 both the second and third roots exceed the .001 level of significance for Alpha.

Table 2 represents just 2 different error structures. Mean root sizes for numerous error structures are shown in Figure 22, and the conclusion is the same: even with minimal nonerror structure built in,the error roots tend to be enlarged.

The error roots, enlarged undesirably with built-in dimensionality, yield, of course, distorted statistics for the error roots. In Figure 23 it is shown that the largest error root tends to be in the critical region when substantial unidimensional structure is built in. The Alpha and Delta statistics for error roots seem to be highly sensitive to unidimensional structure. When multidimensional structure is built in, Beta and Delta statistics for the largest error root remain below the critical level, but they do not rise much higher on the graph for the smallest nonerror root. Gamma seems to be very sensitive to multidimensional structure. From these last results Beta seems to be the most dependable statistic for evaluating the largest error root.

It was pointed out in previous paragraphs that Alpha was sensitive to unidimensional nonerror structure, but apparently raw root sizes are also sensitive to nonerror variance. Variance along a single continuum appears to spill over into many roots. Somehow the MDS computations diffuse the nonerror variance. Error roots are apparently, with nonerror roots, the joint beneficiaries of non-random judging. Although Alpha and Epsilon, both raw root size statistics, are still potentially useful for discerning whether the judgment matrix is based on nothing but random judging; they are apparently of no value in determining the number of nonerror dimensions.

CHAPTER IV

ASCERTAINING DIMENSIONALITY

In almost every real MDS problem there will be some random variation and some nonrandom variation. The investigator will attempt to interpret all the variance associated with roots that are significantly unlike random error roots. In so doing he presumes that some nonerror variance will go undetected. Any root smaller than the smallest nonerror root will be treated as if it represents error variance only. Any root larger than the largest error root will be presumed to involve both nonerror and error variance.

The issue at hand is the identification of the smallest nonerror root. Can the distributions of Chapter II, distributions involving the largest root obtained from completely random input, be used to identify this smallest nonerror root? To answer this question, let us consider the components of variance, error components and nonerror components, separately. Let us assume that with a principal axes representation of the nonerror roots like Figure 24a, the number of nonerror components, k, is much smaller than m-1, the number of roots.

$$k \ll m-1 \tag{24}$$

The m-1 error components would be distributed as in Figure 24b (or Figure 2). Under these conditions it is unlikely that the largest error component would be correlated with any appreciable nonerror component, much less coincide with one.

Another way of saying this is that when dimensionality is imposed on random input the largest nonerror component is unlikely to combine with the largest error component. This situation is illustrated in Figure 24c. Figure 24c should not be interpreted too literally. The nonerror components will probably not combine with any Figure 24b error component, combining rather with some composite of small error components from Figure 24b.

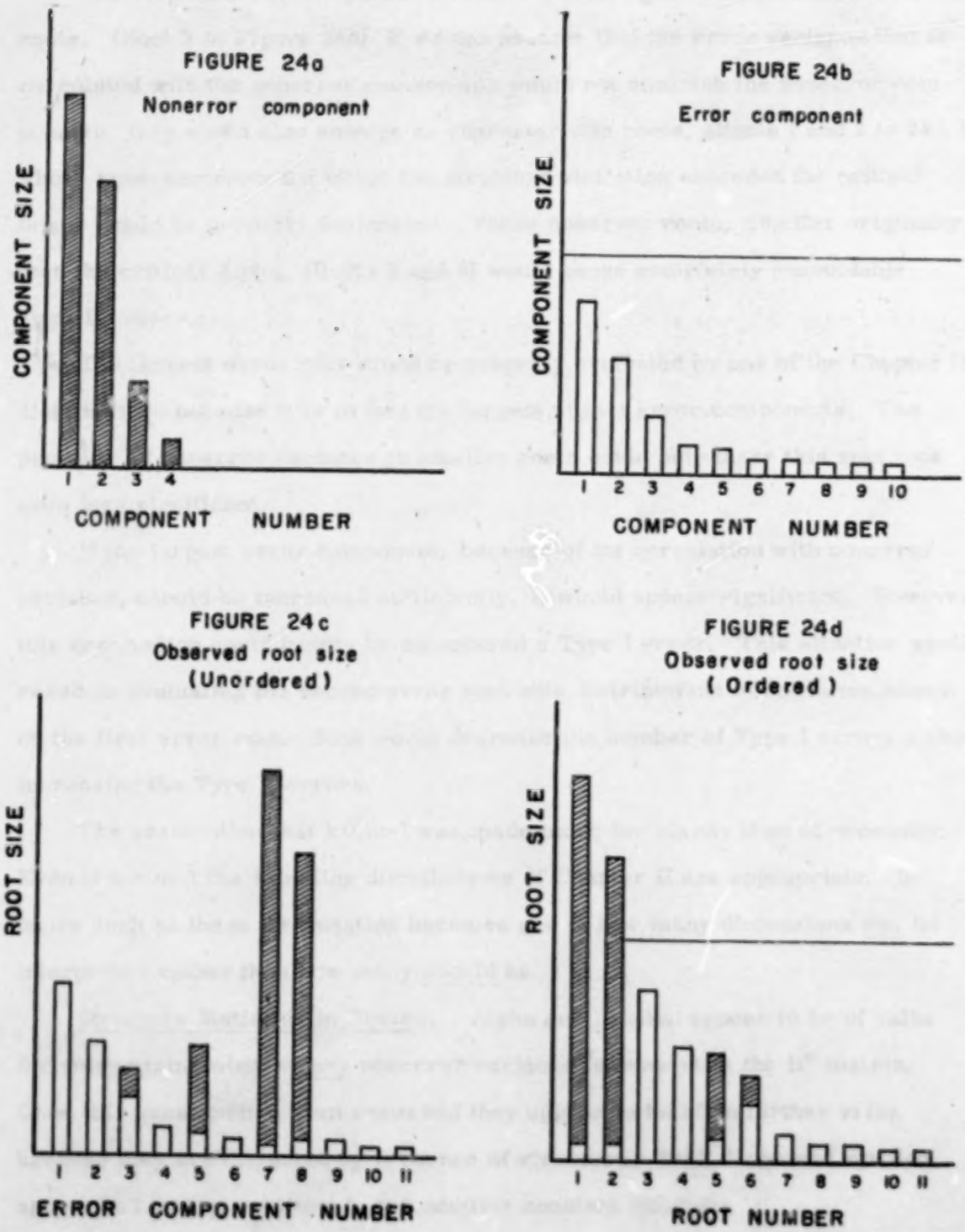


Fig. 24. The components of nonerror variance and error variance alone and in combination.

The largest error component would then emerge as one of the characteristic roots. (Root 3 in Figure 24d) If we can assume that the error variance that is correlated with the nonerror components would not diminish the nonerror components, they would also emerge as characteristic roots, (Roots 1 and 2 in 24d.) Those nonerror roots for which the structure statistics exceeded the critical levels would be properly designated. Those nonerror roots, smaller originally than the critical Alpha, (Roots 5 and 6) would cause completely unavoidable Type II errors.

The largest error root would be properly evaluated by use of the Chapter II distributions because it is in fact the largest of m-1 error components. The pressure of nonerror variance in smaller roots could only make this root look even less significant.

If the largest error component, because of its correlation with nonerror variance, should be increased sufficiently, it would appear significant. However, this designation could hardly be considered a Type I error. This situation would result in evaluating the second error root with distributions of statistics based on the first error root. Such would decrease the number of Type I errors without increasing the Type II errors.

The assumption that $k \ll m-1$ was made more for clarity than of necessity. Even if $k = m-1$ the sampling distributions of Chapter II are appropriate. In cases such as these the question becomes one of how many dimensions <u>can</u> be interpreted rather than how many <u>should</u> be.

<u>Structure Statistics in Review.</u> Alpha and Epsilon appear to be of value for determining whether <u>any</u> nonerror variance is present in the B^* matrix. Once this question has been answered they appear to be of no further value because they are distorted by presence of structure. Both Alpha and Epsilon appear to be very sensitive to the additive constant estimate.

Gamma, like Alpha and Epsilon, tended to be of little value for ascertaining dimensionality. It made both error and nonerror roots look significant. Gamma also was sensitive to the additive constant but in the opposite direction. If both Alpha and Gamma appear significant, then the additive constant estimate is not a source of concern.

Beta and Delta were appropriately sensitive to structure, reflecting nonerror variance in the root under scrutiny and not reflecting nonerror variance in adjacent roots. Small Betas and Deltas can result from nonerror variance in subsequent roots so they should not be computed starting with the largest root and working down but by starting with a clearly indicated error root and working up. Beta appears to perform better when multidimensional nonerror structure is present and Delta when unidimensional nonerror structure is present. Beta and Delta do not appear to be subject to misestimates of the additive constant, and it may be that these two statistics are distributed without regard to number of objects, number of judges and judging procedure followed.

Thus, all five statistics seem to have utility. Since they are all available when the characteristic roots and vectors have been determined they probably all should be calculated and used in evaluating the nonerror variance in an MDS situation.

A Real Live Example. Thirty-five counselor-trainees were asked to consider fifteen members of that group as to similarity as counselors. Multiple rank order procedures were followed. The B^* matrix was obtained and characteristic roots extracted. They were 3.06, 1.33, .97, .80, .60, .47, .38, .15, .06, .02, -.01, -.12, -.27, -.47. Alpha, Gamma, and Epsilon for the largest root were 3.06, .44, and 6.99 respectively. All these values were significant at the .001 level. It was concluded that the comparisons made by these students were not completely random.

If we decide not to consider any root smaller than .50, the .05 critical level for Alpha, we will first consider Root 5 and move back toward Root 1. The structure statistics for Roots 5-1 are shown in Table 3.

TABLE 3

Structure Statistics for Actual MDS Problem

	Root 5	Root 4	Root 3	Root 2	Root 1
Alpha	.61	.81	.98	1.34	3.06
Beta	1.29	1.33	1.21	1.36	2.29
Gamma	.75	.50	.38	.34	.44
Delta	1.44	1.71	1.16	2.14	9.09

Referring to the significance levels reported in Figures 5 and 9, it is apparent that Betas and Deltas of only the largest root are statistically significant. All other Beta and Delta values are too small to be distinguished from error roots. The Alphas and Gammas for Roots 5-2 are significant according to Figures 3 and 7, but this must be discounted because it could be due to large unidimensional variance. The tendency of Gamma to increase in the smaller roots is apparent.

Artificial Structure Approximations of the Actual Data. Many of the runs with artificially built-in structure yielded statistics similar to the actual data run just mentioned. It was therefore a routine matter to examine those runs to see what artificial structure most nearly approximated the actual data structure. In Figure 25 the actual data run is designated Run X. The artificial structure runs are identified by a run number.

With one dimension built in, the curves broke abruptly at the second root, much more so than Run X. A unidimensional run with AIDP of .4 would have given about the right size Root 1 but its Roots 2-7 would have been considerably smaller than on Run X.

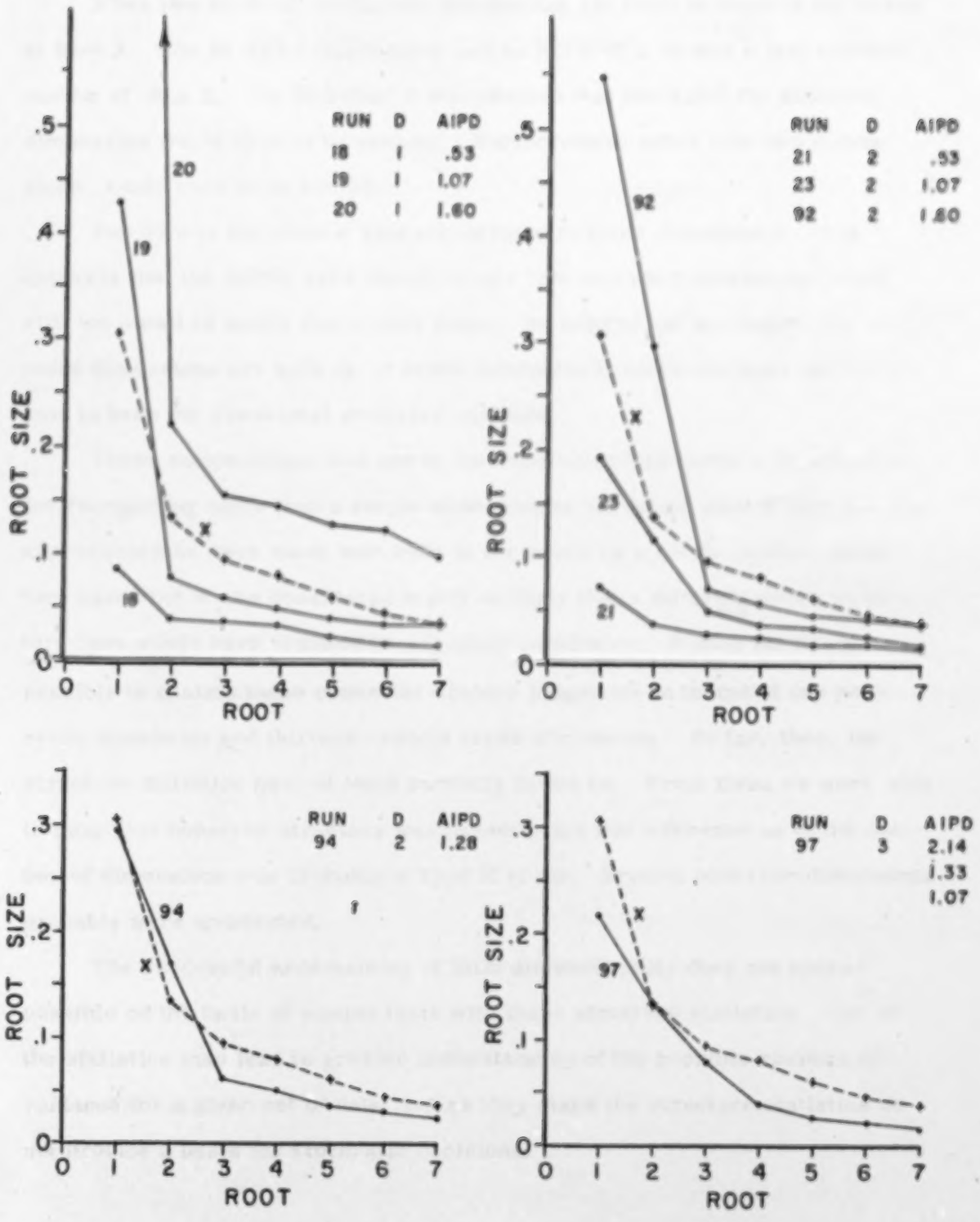


Fig. 25. Approximations to actual-data MDS Run X by runs with artificially built-in structure.

When two or three substantial dimensions are built in there is no "break" at Root 2. Run 94 with 2 dimensions and an AIPD of 1.28 was a fair approximation of Run X. To do better it was obvious that the AIPD for different dimensions would have to be unequal. Furthermore, more than two dimensions would have to be loaded.

Run 97 was the closest approximation with three dimensions. It is apparent that the AIPD, even though larger than any used previously, were still too small to match Run X root sizes. As pointed out in Chapter III, as more dimensions are built in, it takes substantially more variance per dimension to keep the same total structure variance.

These comparisons lead one to the conclusion that there is justification for recognizing more than a single dimension in the actual data of Run X. The approximations were made with built-in structure on a single random structure base, but it was considered highly unlikely that a different error structure base would have resulted in any other conclusion. It does not seem sensible to explain these counselor-trainee judgments in terms of one nonerror dimension and thirteen random error dimensions. So far, then, the structure statistics have at least partially failed us. From them we were able to infer that nonerror structure was present, but our inference as to the number of dimensions was probably a Type II error. Several nonerror dimensions probably were undetected.

The successful ascertaining of MDS dimensionality does not appear possible on the basis of simple tests with these structure statistics. Use of the statistics may lead to greater understanding of the possible sources of variance for a given set of data, but as they stand the structure statistics do not provide a basis for stochastic decisions.

CHAPTER V

SUMMARY

MDS and Educational Measurement. Measurement is generally accepted by educators as being essential for the effective description and evaluation of educational phenomena. Educational measurement is dependent not only upon the construction of reliable scales and the expert manipulation of these scales, but also upon adequate development of interpretive procedures. It is the refinement of one of these procedures, multidimensional scaling (MDS), with which this research was concerned.

The complexities of human behavior appear infinite, but the lawfulness of its patterns are readily apparent. That the uniqueness of personal traits and the seeming capriciousness of individual acts can be reduced to finite, orderly description has always been the major premise of the social scientist. The same premise was expanded at length by Thurstone who suggested the use of the spatial models for the description of psychological variables and who, with numerous others, developed the mathematical steps appropriate to this reduction (1927, 1947).

Multidimensional scaling is a psychometric reduction process based upon a spatial model. It is used to identify the "space" or "domain" defined by the critical characteristics of a given set of objects. It was first introduced by Richardson (1938) and developed through the contributions of Young and Householder (1938), Klingberg (1941), Attneave (1949), and Torgerson (1952). Recent writings by Messick (1956) and Torgerson (1958) reveal the current state of development of this scaling methodology.

Multidimensional scaling, although not limited to analyses of educational data, shows promise of much future value in many important educational

problems, e. g., the definition of the roles of specialists such as the counselors or team teaching coordinators; the determination of the dimensions of academic achievements; and the investigation of the "structure" of a group of learners. MDS procedure involves multiple assessments of the similarities among a set of stimulus objects. It is assumed that these objects may be similar in different ways and that the assessments may embody these different similarities. Analysis of these assessments through MDS leads to identification of the critical factors or dimensions along which the objects were similar or dissimilar. The identification of critical factors or underlying dimensions is a primary objective in any major scientific undertaking.

Theoretically, the size of error factors could be obtained through mathematical statistics. A mathematical solution, however, is not expected because the problem is too complicated and too trivial, i. e., lacking mathematical significance. It is the opinion of the mathematical statistician that an investigation of the error factors in MDS will occur only if methodologists undertake empirical studies. This study is one such empirical study.

Objectives. The purpose of this research was to generate the sampling distributions necessary for discriminating between error and nonerror MDS dimensions. By determining the likelihood that characteristic roots of a given size are error roots, statistical tables were to be developed for use in testing the null hypothesis: A given MDS factor is no greater than one attributable to a combination of measurement errors.

It was recognized, however, that the sampling distributions of error roots might be influenced by certain variables such as (a) the number of stimulus objects assessed or judged, (b) the number of judges, (c) the type of stimulus arrangement, e. g., complete method of triads, multidimensional

ranking, (d) the type of factor analysis used in the analysis, and (e) the pattern of nonerror factors present. These variables were considered after the first set of sampling distributions was determined.

Procedure. The procedure for this research had three phases: (a) development of the computer programs, (b) generation of the sampling distributions, and (c) investigation of the several variables mentioned in the preceding paragraph. The computer programs were based upon the procedure outlined in Torgerson (1958). The characteristic roots of the scalar product matrix were the source of five structure statistics.

Initial activity involved the simulation of judgments of 15 objects by 35 judges according to the multidimensional ranking procedure (Klingberg, 1941). The effect of judge number was varied,as the first step in the third phase, to 20 and to 80 judges. The number of objects was varied independently to 8 as well as 15. The difference in using a complete method of triads stimulus arrangement was considered.

The Structure Statistic. Alpha, the raw size of a root, represented the variance associated with a dimension of the MDS structure. When based entirely on random input, the distribution of this statistic appeared to be dependent on the number of stimulus objects and judges, and the scaling method used. Alpha varied considerably with different estimates of the additive constant. When structure was built in, the variance of error roots was undesirably increased. Alpha does appear to be of some value in determining whether or not any nonerror variance is present in the judges' responses, but appears to have no utility for ascertaining the dimensionality.

Beta, the ratio of the size of a root to the size of the next smaller root, is a relative index of variance associated with a dimension. Beta apparently

does not covary with the number of judges used but does increase as the number of stimulus objects is decreased. Errors in the estimation of the additive constant had only a small effect on Beta. Betas for error dimensions do not appear to be distorted when large amounts of nonerror variance are associated with other dimensions, but they may not be sensitive enough to detect all the nonerror variance in a matrix that should be detected.

Gamma, the proportion of total variance associated with a given dimension, performed pretty much as Alpha did. It, however, was sensitive to the additive constant in an opposite way, making possible the joint use of Alpha and Gamma to discount the importance of estimation of this constant.

Delta, the ratio of the difference between two roots and an average difference, was similar to Beta in all important statistical respects.

Epsilon, an index of the total variance, was seen, like Alpha, to be useful only in determining the presence of any nonerror variance in the entire judgment matrix.

Final Conclusions. The size of the roots of an MDS B* matrix obtained from completely random multiple rank order data is a function of the number of objects and number of judges. The more judges, the smaller the roots. The more numerous the objects, the larger the roots. In any case about 1/3 of the roots are likely to be negative.

The mean root sizes were shown for 100 runs with 15 objects and 35 judges. The line connecting these roots is almost rectilinear with slightly more variance associated with the terminal roots. At first it was thought that the departure from linearity might be attributable to improper estimation of the additive constant, but the same enhancement of first roots and last roots is apparent regardless of the size of the additive constant. The plot of root sizes for any single random-data run is rather irregular. When unidi-

mensional structure is present, the roots size line becomes much less regular even for the middle roots. Visual scrutiny of the plot to determine dimensionality should be undertaken only with considerable caution.

Use of the structure statistics was demonstrated with reference to an actual MDS problem. The procedures seemed to lead to a satisfactory conclusion, i.e., that only a single nonerror dimension could be designated with statistical confidence. In an attempt to approximate the observed structure with artificial structure added to error structure, it was found that more than a single dimension, probably four or five dimensions, were required. It was concluded, therefore, that ascertaining the nonerror dimensionality of the structure could not be accomplished with these structure statistics alone. The authors concluded, however, that there is sufficient systematic relationship between the presence of nonerror structure and the pattern of characteristic roots to justify additional investigation along these lines.

REFERENCES

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley and Sons, 1958.

Attneave, F. A method of graded dichotomies for the scaling of judgments, Psychol. Rev., 1949.

Brownlee, J. Log (X) from X = 1 to 50.9 by intervals of .01. London: Cambridge University Press, 1923.

Gulliksen, H. Linear and multidimensional scaling, Psychometrika, 1961, 26, 9-25.

Hefner, R. A. Extensions of the law of comparative judgment to discriminal and multidimensional stimuli. Unpublished doctoral dissertation, University of Michigan, 1958.

Helm, C. A multidimensional ratio scaling analysis of perceived color relations. ONR Technical Report. Princeton University, 1959.

Kendall, Maurice G. The advanced theory of statistics. Vol. 1. London: J. B. Lippincott, 1943.

Klingberg, F. L. Studies in measurement of the relations among sovereign states, Psychometrika, 1941, 6, 335-352.

Mallows, C. L. Latent roots and vectors of random matrices. Princeton, N. J.: Statistical Techniques Research Group, Technical Report 35, 1961.

Messick, J. S. Some recent theoretical developments in multi-dimensional scaling, Educ. Psychol. Measmt., 1956, 16, 82-100.

Messick, S. J. The perception of social attitudes, J. Abnorm. Soc. Psychol., 1956, 52, 57-66.

Messick, S. J. and Abelson, R. P. The additive constant problem in multi-dimensional scaling, Psychometrika, 1956, 21, 1-17.

Morton, A. S. Similarity as a determinant of friendship: a multi-dimensional study. Princeton University Psychology Department, 1959.

Pairman, E. Tables of the digamma and trigamma function. London: Cambridge University Press, 1919.

Pearson, E. S. Table of the logarithms of the complete gamma function. London: Cambridge University Press, 1922.

Pearson, K. Tables of the incomplete beta function. London: Cambridge University Press, 1934.

Pearson, K. Tables of the incomplete gamma function. London: Cambridge University Press, 1951.

Richardson, M. W. Multidimensional Psychophysics, Psychol. Bull., 1938, 35, 659-660.

Shepard, R. N. The analysis of proximities: multidimensional scaling with an unknown distance function, Bell Telephone Lab, Murray Hill, New Jersey: author, 1961.

Sjogren, D. D. Changes in the organization of mental ability among secondary school youth. Unpublished doctoral dissertation, University of Nebraska, 1961.

Torgerson, W. S. Multidimensional scaling: I. theory and method, Psychometrika, 1952, 17, 401-419.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley and Sons, 1958.

Thurstone, L. L. A law of comparative judgment, Psychol. Rev., 1927, 34, 273-286.

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Young, G. and Householder, A. S. Discussion of a set of points in terms of their mutual distances, Psychometrika, 1938, 3, 19-22.